IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application No. 10/513,961

Confirmation No. 4710

Applicant: Boyd et al.

Filed: December 20, 2004

TC/AU: 1648

Examiner: Zachariah Lucas

Docket No.: 231119 (Client Reference No. E-017-2002/0-US-06)

Customer No.: 45733

DECLARATION UNDER 37 C.F.R. § 1.132 OF BARRY R. O'KEEFE, PH.D.

I, Barry R. O'Keefe, Ph.D., do hereby declare:

1.   I am a co-inventor of the subject matter disclosed and claimed in the above-captioned patent application.

Scytovirin Fragments and Variants

2.   The amino acid sequence of wild-type scytovirin is described in the present application (i.e., SEQ ID NO: 1, which is 95 amino acids in length).

3.   A DNA construct encoding amino acids 1-48 of scytovirin (SD1), wherein the cysteine at residue 7 is substituted with serine, was constructed, expressed, and isolated. An XTT-tetrazolium based assay was used to determine the anti-HIV activity of SD1 and scytovirin against acute HIV-1 infection in CEM-SS cells. The protocol followed was the same as described in the present application (see, e.g., Example 5).

4.   SD1 and scytovirin showed comparable activity against the T-tropic laboratory strain HIV-1$_{RF}$ in CEM-SS cells with EC$_{50}$ values of 6.6 nM and 7.5 nM, respectively. Toxicity to the CEM-SS cell line was not detected for either scytovirin or SD1 at concentrations up to 10,000 nM. Thus, the SD1 fragment of scytovirin retained scytovirin's potent anti-viral (e.g., anti-HIV) activity.

5.   Additional experiments were performed to determine whether the N- and C-terminus of SD1 are necessary for antiviral activity. In particular, a series of deletion mutants were constructed in which two, five, or ten amino acids were deleted from the N-terminus of SD1, or eight amino acids were deleted from the C-terminus of SD1 as described in Xiong et al., *Peptides, 27*: 1668-1675 (2006) (Attachment A). The deletion mutants were subjected to the XTT-tetrazolium based assay.

6.   Deletion of two, five, or ten N-terminal amino acids from SD1 completely eliminated antiviral activity, indicating that N-terminal amino acids of SD1 are necessary for maintaining the antiviral activity of SD1. Deletion of the eight amino C-terminal amino acids resulted in a 3- to 7-fold decrease in anti-HIV potency, indicating that these C-terminal amino acids optimize the activity of SD1, but are not necessary in order to retain some anti-viral activity.

7.   In addition to the scytovirin fragments and variants described above, it is possible to generate proteins comprising several mutations in the scytovirin amino acid sequence (e.g., proteins comprising 90% identity to the amino acid sequence set forth in SEQ ID NO: 1), which retain antiviral activity (see paragraphs [0025]-[0026] of the specification). Moreover, one of ordinary skill in the art has the ability to determine which amino acids of SEQ ID NO: 1 to alter to generate mutant scytovirins having antiviral activity using routine laboratory techniques.

8.   To identify amino acid residues appropriate for manipulation to generate a functional scytovirin protein, the ordinarily skilled artisan can determine the three-dimensional structure of the scytovirin protein from SEQ ID NO: 1. Indeed, the three-dimensional structure of scytovirin is described in McFeeters et al., *J. Mol., Biol., 369*: 451-461 (Attachment B). Ideally, mutations that do not modify the electronic or structural environment of the protein are generated to retain optimal antiviral activity. By utilizing information regarding the three-dimensional structure of the protein and the amino acid sequence of SEQ ID NO: 1, determination of which amino acids are critical for proper protein folding by way of their location within the protein structure or interaction with surrounding residues is within the skill of the ordinary researcher.

9.   In addition to creating mutations in regions of the amino acid sequence of SEQ ID NO: 1 that are not critical for three-dimensional structure, the ordinary

researcher can determine which amino acid residues are likely responsible for viral binding. It is understood in the art that surface hydrophobicity plays a key role in protein-protein interactions, such as the interaction between scytovirin and viral proteins. The ordinary researcher has the ability to map hydrophobic surface clusters on the scytovirin protein or scytovirin homologs to predict regions critical for interaction with the viral envelope using routine methods such as those disclosed in McFeeters et al., *supra*; Botos et al., *J. Biol. Chem., 277(13)*: 34336-34342 (2002) (Attachment C); and Bewley et al., *Nature Structural Biology, 5(7)*: 571-578 (1998) (Attachment D). Amino acid residues not found in hydrophobic surface clusters are likely not critical for hydrophobicity of these clusters and, thus, are appropriate targets for mutation to create scytovirin variants (e.g., with 90% identity to SEQ ID NO: 1), which retain antiviral activity.

10.   With respect to screening for antiviral activity, the assays described in the present application (see, e.g., Example 5) are well within the skill of the ordinary researcher and require only routine laboratory techniques.

Scytovirin Antiviral Activity

11.   As described in the specification, scytovirin, as well as variants and fragments thereof, have potent antiviral activity. In addition to anti-HIV activity, scytovirin has potent anti-viral activity against Ebola and influenza viruses.

12.   To demonstrate scytovirin's activity against Ebola virus, VeroE6 cells plated in 96 well plates were infected with recombinant Ebola (Zaire) virus expressing green fluorescent protein (GFP). Various concentrations of scytovirin and cyanovirin were introduced to the cells before Ebola virus infection. After incubation for 48 hours, GFP expression was detected by fluorescence.

13.   Scytovirin demonstrated potent antiviral activity with an EC_{50} value of about 0.4 µg/ml, while cyanovirin had an EC_{50} value of 3.5 µg/ml. When greater than about 1 µg/µl scytovirin was added to the cells before Ebola virus infection, almost no GFP expression was detected, indicating scytovirin's inhibition of Ebola virus infection. In comparison, even at concentrations of about 100 µg/µl cyanovirin, fluorescence was detected, indicating infection of the cells by Ebola virus.

14.   To further demonstrate scytovirin's activity against Ebola virus, 20 g female Balb/C mice were challenged with 1000 pfu/mouse of mouse-adapted Ebola

(Zaire) virus. Twenty-four hours prior to viral challenge, the mice began receiving twice-daily (Q12) 10 mg/kg subcutaneous injections of scytovirin in phosphate buffered saline (PBS).

15.   Due to the highly infectious nature of Ebola virus, the ability of scytovirin to inhibit viral infection in mice at risk thereof was evaluated by prolonged survival of treated mice compared to untreated control mice. By Day 8, only 20% of the control mice survived. In contrast, 60% of the scytovirin-treated mice survived. Indeed, 60% of the scytovirin-treated animals survived until Day 20, the last timepoint of the study.

16.   Thus, scytovirin has potent antiviral activity against Ebola virus both *in vitro* and *in vivo*.

17.   The anti-influenza virus activity of scytovirin was tested in the following manner. Host cells used for these assays were obtained at the Southern Research Institute-Frederick (SoRI-Frederick, Frederick, Maryland). MDCK cells were used for all assays. The humanized H5N1 influenza virus strains that were used in the assays included Hong Kong/491/1997 and Vietnam/1203/2004H. The avian H5N1 influenza virus strains that were used in the assays included Duck/MN/1525/81 and Gull/PA/4175/83. Additionally, humanized influenza virus strains B/Shanghai/362/02, A/New Caledonia/20/90 (H1N1), and A/California/7/04 (H3N2) were used.

18.   A typical antiviral assay for each virus was as follows. A pretreated aliquot of virus was removed from the freezer (-80° C) and allowed to thaw. The virus was diluted into tissue culture medium such that the amount of virus added to each well in a volume of 100 μl was that amount pre-determined to yield complete cell killing at 3-7 days (depending on the virus) post-infection. Scytovirin (SEQ ID NO: 1) stock solution was diluted into the medium to the desired high concentration and then further diluted in the medium in 0.5 log10 increments.

19.   The day following plating of cells, plates were removed from the incubator, and medium was removed and discarded. The infection medium was DMEM (Dulbecco's Minimal Essential Medium) supplemented with 0.3% BSA, 100 U/ml penicillin, 100 μg/ml streptomycin, 0.1 mM non-essential amino acids, 0.1 mM sodium pyruvate, 2 mM L-glutamine, and 2.0 μg/ml TPCK-treated (N-tosyl-L-

phenylalanine ketone) trypsin. Drug dilutions (6-12 dilutions) were made in medium and added to appropriate wells of a 96-well microtiter plate in a volume of 100 µl per well. Each dilution was set up in triplicate. Infection medium containing appropriately diluted virus was added to appropriate wells of the microtiter plate. Each plate contained cell control wells (cells only), virus control wells (cells plus virus), drug toxicity control wells (cells plus drug only), drug colorimetric control wells (drug only), as well as experimental wells (drug plus cells plus virus).

20. After 7 days of incubation at 37° C in a 5% CO_2 incubator, the test plates were analyzed by staining with CellTiter 96^{TM} AQ$_{ueous}$ One Solution Reagent (Promega, Madison, WI). Twenty microliters of the reagent were added to each well of the plate, and the plate was reincubated for 4 hrs at 37° C. CellTiter 96^{TM} AQ$_{ueous}$ One Solution Reagent contains a tetrazolium (MTS) compound and the electron-coupling reagent phenazine ethosulfate (PES). MTS-tetrazolium is bioreduced by cells into a colored formazan product that is soluble in tissue culture medium. This conversion is mediated by NADPH and NADH produced by dehydrogenase enzymes in metabolically active cells. The amount of soluble formazan produced by cellular reduction of the MTS was determined by measuring the absorbance of light at a wavelength of 490 nm. The quantity of formazan product as determined by the amount of 490 nm wavelength light absorbance was directly proportional to the number of living cells in the culture, thus allowing the rapid quantitative analysis of the inhibition of virus-induced cell killing by the test substances.

21. The IC_{50} (50% inhibition of virus replication), EC_{50} (50% reduction in cell viability), EC_{90} (90% reduction in cell viability), and the corresponding antiviral index (IC_{50}/EC_{50}) were calculated using standard procedures. Results from testing of scytovirin (SEQ ID NO: 1) against different strains of influenza virus are shown in the tables below.

Table 1: Anti-H5N1 Influenza Virus Activity of Scytovirin

Virus Strain	Assay	EC_{50} (µg/ml)	EC_{90} (µg/ml)	IC_{50} (µg/ml)	Antiviral Index
Duck/MN/1525/81	Visual	20		>100	>5
	Virus Yield		>100	>100	
Gull/PA/4175/83	Visual	20		>100	>5
	Virus Yield		33	>100	>3
Hong Kong/491/1997	Visual	0.4		>100	>250
	Virus Yield		>100	>100	
Vietnam/1203/2004I1	Visual	4		>100	>25
	Virus Yield		2	>100	>50

Table 2: Anti Influenza Virus Activity of Scytovirin

Virus Strain	EC_{50} (µg/ml)	IC_{50} µg/ml
B/Shanghai/362/02	0.18	>10
A/New Caledonia/20/99 (H1N1)	0.14	>10
A/California/7/04 (H3N2)	0.18	>10

22.   As illustrated by the data set forth in the table, scytovirin is active against all tested strains of humanized influenza virus as demonstrated by an EC_{50} value of less than 10 µg/ml. Accordingly, scytovirin is effective for inhibiting an influenza viral infection of a host, and in particular humanized influenza viral infections.

23.   I hereby declare that all statements made herein of my own knowledge are true, that all statements made on information and belief are believed to be true, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Date: July 30, 2007
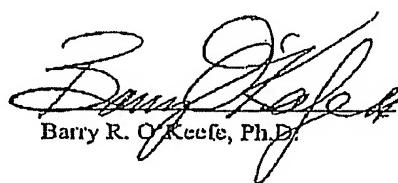
Barry R. O'Keefe, Ph.D.

6

Attachment A
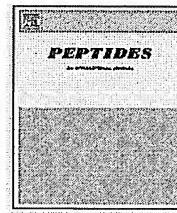
Potent anti-HIV activity of scytovirin domain 1 peptide

Changyun Xiong [a], Barry R. O'Keefe [a,*], R. Andrew Byrd [b], James B. McMahon [a]

[a] Molecular Targets Development Program, National Cancer Institute-Frederick, National Institutes of Health,
Bldg 562/Rm 201, Fort Detrick, Frederick, MD 21702, USA
[b] Structural Biophysics Laboratory, National Cancer Institute-Frederick, National Institutes of Health, Frederick, MD, USA

ARTICLE INFO

ABSTRACT

Scytovirin (SVN) is a novel anti-HIV protein isolated from aqueous extracts of the cultured cyanobacterium *Scytonema varium*. SVN contains two apparent domains, one comprising amino acids 1–48 and the second stretching from amino acids 49 to 95. These two domains display significant homology to each other and a similar pattern of disulfide bonds. Two DNA constructs encoding scytovirin 1–48 (Cys7Ser) (SD1) and 49–95 (Cys55Ser) (SD2) were constructed, and expressed in *E. coli*, with thioredoxin fused to their N-terminus. Purified recombinant products were tested for binding activities with the HIV surface envelope glycoproteins gp120 and gp41. Whole cell anti-HIV data showed that SD1 had similar anti-HIV activity to the full-length SVN, whereas SD2 had significantly less anti-HIV activity. Further deletion mutants of the SD1 domain (SVN(3–45)Cys7Ser, SVN(6–45)Cys7Ser, SVN(11–45)Cys7Ser) showed that the N-terminal residues are necessary for full anti-HIV activity of SD1 and that an eight amino acid deletion from the C-terminus (SVN(1–40)Cys7Ser) had a significant effect, decreasing the anti-HIV activity of SD1 by approximately five-fold.

Published by Elsevier Inc.

1. Introduction

Viral infections remain among the most formidable causes of human and non-human animal morbidity and mortality worldwide. Effective prevention and therapies against most viral pathogens remain elusive. One of the most recent and catastrophic examples is the rapidly expanding and pervasive worldwide pandemic of human immunodeficiency virus (HIV) infection and acquired immune deficiency syndrome (AIDS). The need for effective preventative and therapeutic agents for HIV/AIDS and other potentially lethal viral diseases remains an urgent global priority.

Antimicrobial peptides and proteins are the essential molecular effectors of innate immunity in human and other vertebrates to combat microbial challenge [7,8]. Peptides offer tremendous structural diversity that can be exploited for the development of novel therapeutics and prophylactics for

many different diseases. For example, in the field of HIV therapeutics a novel, rationally constructed peptide molecule known as T-20 [12] has been shown to be a potent inhibitor of HIV/cell fusion. Furthermore, naturally occurring, non-mammalian peptides and proteins offer new avenues for antiviral discovery and development [4,5,17].

Scytovirin (SVN), a potent anti-HIV protein, was originally isolated from aqueous extracts of the cyanobacterium *Scytonema varium* [3]. SVN, with a molecular mass of 9713 Da, contains five intrachain disulfide bonds and binds to the HIV-1 envelope glycoproteins gp120, gp160 and gp41, but not to the cellular receptor CD4 [3]. Low nanomolar concentrations of SVN inactivate both laboratory strains and primary isolates of HIV-1. This inhibition has been shown to involve selective interactions between SVN and high-mannose oligosaccharides [3]. Specifically, SVN interacts with oligosaccharides containing $\alpha 1$–2, $\alpha 1$–2, $\alpha 1$–6 tetramannoside units [2].

```
1   GSGPTYCWNEANNPGGPNRCSNNKQCDGARTCSSSGFCQGTSRKPDPG  48
    ======*==  ==========*====================*
49  PKGPTYCWDEAKNPGGPNRCSNSKQCDGARTCSSSGFCQGTAGHAAA  95
```
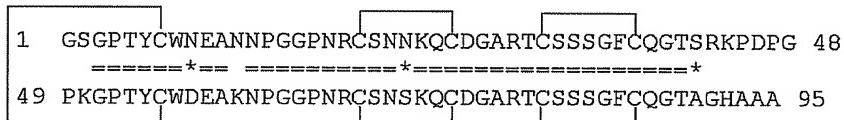
Fig. 1 – Sequence alignment between the two domains of SVN. Amino acids are represented by single-letter codes. Identical sites are represented by '='. Conserved sites are represented by '*'. Disulfide bonds are marked with solid lines above the sequence.

SVN shows strong internal sequence duplication. When amino acids 1–48 and 49–95 are aligned, 36 residues (75%) are identical and three (6%) represent conservative amino acid changes (Fig. 1). Furthermore, NMR analysis of SVN revealed that the protein has two domains with strong symmetry [Drs. McFeeters and Byrd, personal communication]. Based on the above facts, DNA constructs for SVN single domains [amino acids 1–48 (SD1) and 49–95 (SD2)] were produced and expressed to explore functional domains and sequence requirements for gp120 (or gp41) binding and anti-HIV activity of SVN. Furthermore, additional truncation mutants of SD1 were expressed and their anti-HIV activity assessed.

2. Materials and methods

2.1. Materials

Restriction endonucleases and T4 DNA ligase were obtained from New England Biolabs (Beverly, MA). Pfu DNA polymerase was obtained from Stratagene (La Jolla, CA). Rabbit anti-SVN polyclonal antibodies were produced as described previously [3]. Recombinant enterokinase was from Novagen (Madison, WI). Recombinant glycosylated gp120 was obtained from Intracel Corporation (Cambridge, MA). Glycosylated gp41 (HIV-1 gp41hp) was purchased from Viral Therapeutics, Inc. (Ithaca, NY). Reagents for sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS-PAGE) were obtained from Invitrogen (Carlsbad, CA). All other chemicals were analytical reagent grade from Sigma (St. Louis, MO). The plasmid vector pET32C(+) and E. coli strain Origami B(DE3)pLysS were from Novagen (Madison, WI).

2.2. DNA amplification

For PCR amplification, pET(SVN) plasmid [21] was used as a template. Primers synthesized by Integrated DNA Technologies, Inc. (Coralville, IA) are as follows:

SD1 (SVN(1–48)Cys7Ser):
Forward primer: 5'-CATGCCATGGCTGGTTCTGGTCCGACC-TACTCTTGGAACG-3'
Reverse primer: 5'-GCGCTCGAGTTACCCCGGGTCCGGTT-TACGAGA-3'
SD2 (SVN(49–95)Cys55Ser):
Forward primer: 5'-CATGCCATGGCTCCCAAAGGTCCGAC-CTACTCTTGGGACGAG-3'
Reverse primer: 5'-CCCGCTCGAGTTACGCAGCCGCGTGAC-CCGCGG-3'

SVN(3–45)Cys7Ser:
Forward primer: 5'-CATGCCATGGCTGGTCCGACCTACTC-TTGGAAC-3'
Reverse primer: 5'-GCGCTCGAGTTACGGTTTACGAGAGG-TACCCTGGC-3'
SVN(6–45)Cys7Ser:
Forward primer: 5'-CATGCCATGGCTTACTCTTGGAACGA-AGCGAAC-3'
Reverse primer: 5'-GCGCTCGAGTTACGGTTTACGAGAGG-TACCCTGGC-3'
SVN(11–45)Cys7Ser:
Forward primer: 5'-CATGCCATGGCTGCGAACAACCCGGG-TGGTCCGAAC-3'
Reverse primer: 5'-GCGCTCGAGTTACGGTTTACGAGAGG-TACCCTGGC-3'
SVN(1–40)Cys7Ser:
Forward primer: 5'-CATGCCATGGCTGGTTCTGGTCCGACC-TACTCTTGGAACG-3'
Reverse primer: 5'-GGTTTCTGCCAGGGTTAATCTCGTAA-ACCGGAC-3'

The above PCR fragments were digested with NcoI-XhoI, respectively, and inserted into a NcoI-XhoI-digested pET32C(+) vector. The linker sequence between the enterokinase digestion site and SVN DNA was deleted from the construct by site-directed mutagenesis (GeneTailor Site-Directed Mutagenesis System, Invitrogen) and the resulting plasmid vector was transformed into E. coli strain DH5α. The sequence of the gene was confirmed by DNA sequencing.

2.3. Expression of SVN fusion protein in E. coli

The above plasmids expressing various SVN mutants under the control of a T7 promoter were transformed into E. coli Origami B(DE3)pLysS. The cells containing the constructs were grown at 37 °C in LB medium (Luria–Bertani broth) containing 100 μg/mL ampicillin, 34 μg/mL chloramphenicol and 15 μg/mL kanamycin to an OD_{600} of 0.5–0.8. The cells were induced with isopropyl-1-thio-β-D-galactopyranoside (IPTG) to a final concentration of 1 mM, then incubated for an additional 4 h and collected by centrifugation. The cells were either prepared immediately for SDS-PAGE assay or frozen at −80 °C.

2.4. Cell fractionation

Frozen cells were thawed in 50 mM phosphate buffer containing 0.4 mg/mL lysozyme and 1 mM phenylmethylsulfonylfluoride (PMSF). Efficient lysis occurred over several minutes. DNase I (20 μg/mL) and $MgCl_2$ (10 mM) were then added to

digest DNA. Crude lysates were next centrifuged to separate soluble and insoluble protein fractions.

2.5. Peptide purification

The expressed peptides were initially purified by histidine tag metal affinity chromatography (BD) according to the manufacturer's protocol. Briefly, 25 mL cleared lysate from 1 L of LB cultured bacterial pellet was incubated with 2 mL affinity resin that had been equilibrated with 10 column volumes of buffer A (50 mM phosphate buffer, pH7.5, 300 mM NaCl, 5 mM imidazole, 1 mM PMSF). The protein-bound resin was then washed with 10 column volumes of buffer A. SVN mutant fusion proteins were then eluted with 10 column volumes of 150 mM imidazole in buffer A.

The resulting samples were filtered through an Amicon Centriprep 3 kDa molecular mass cutoff filter to a volume of 2 mL to desalt and concentrate the material. They were then digested for 16 h at room temperature with recombinant enterokinase (rEK, Novagen), at 1 unit of enterokinase/10 µg of recombinant protein. rEK-cleaved SVN mutants were then purified using reversed-phase HPLC. The peptides were loaded on a Dynamax C18 300 Å column and eluted with a gradient of 0 to 60% acetonitrile in 0.05% aqueous TFA over 60 min at a flow rate of 3 mL/min with UV monitoring at 210 and 280 nm.

2.6. Mass spectroscopic analysis

The purified recombinant SVN mutants were analyzed by LC–MS using a Agilent high-performance liquid chromatography/electrospray ionization quadrupole mass spectrometer (model 1100D) using a C8 Zorbax column (2.1 mm × 110 mm) eluted with a linear gradient from 0 to 100% acetonitrile over 35 min in H_2O with 5% (v/v) acetic acid in the mobile phase at a flow rate of 0.2 mL/min and with UV monitoring at 280 nm. After mass spectral deconvolution according to manufacturer's protocols, the masses of the various SVN mutants were calculated.

2.7. Immunobloting

Peptide samples were separated by Bis–Tris SDS/PAGE using a 10% (w/v) polyacrylamide gel and electroblotted on to a PVDF membrane. For Coomassie staining 3 µg/well of SV-N and 700 ng/well of SD1 or SD2 were electrophoresed. For Western blotting, 500 ng/well SVN and 200 ng/well of both SD1 and SD2 were used. After blocking in 10mM Tris/HCl, pH 8.0, containing 150 mM NaCl, 0.05% Tween 20 and 5% (w/v) non-fat milk for 1 h at room temperature, the membrane was incubated with rabbit anti-SVN polyclonal antibody at 1:5000 dilution overnight at 4 °C. The immune complexes on the membrane were then reacted with horseradish peroxidase-conjugated goat anti-rabbit IgG at 1:2000 dilution for 1 h at room temperature. Immunodetection was achieved by enhanced chemiluminescence (ECL) according to the manufacturer's protocols. Densitometric scanning of ECL-blots was performed on a Molecular Dynamics 300S computing densitometer (Sunnyvale, CA) using ImageQuant V3.0 software.

2.8. Disulfide bond determination

The presence and connectivity of disulfide bonds were determined as previously described [3]. Briefly, a 100 µg sample of recombinant, non-reduced SD1 was added to 60 µL of 100 mM ammonium bicarbonate (pH 8.0), 6 µL of acetonitrile and 0.6 µL of a 40 µM solution of trypsin in H_2O. The mixture was incubated at 37 °C for 16 h and then analyzed by LC–MS as above. The resulting peptides were evaluated using peptide recognition software to detect all possible disulfide-linked peptide fragments [1].

2.9. Anti-HIV assays

An XTT-tetrazolium-based assay was used to determine the anti-HIV activity of SVN derived peptides on acute HIV-1 infection in CEM-SS cells as previously described [9,16]. Eight serial 1/2 log dilutions of SVN mutants in complete medium were performed and the peptide solutions were then added to exponentially growing CEM-SS human lymphocyte cells. Cell cultures were infected with freshly thawed solutions of HIV-1_{RF} and allowed to incubate for 7 days. Metabolic reduction of the tetrazolium salt, XTT, to a colored formazan product was used to determine cellular viability at the end of the 7-day incubation period.

2.10. ELISA protocols

To determine the binding of SVN derived mutants to glycosylated gp120 and gp41, the envelope glycoproteins were evaluated by ELISA as previously described [10,18]. Briefly, 100 ng/well of either glycoprotein was bound to a 96-well plate, which was then rinsed three times with PBS containing 0.05% Tween 20 (TPBS), and blocked with BSA. Between subsequent steps, the plate was again rinsed with TPBS (×3). The wells were incubated with serial dilutions of each mutant peptide, followed by incubation with a 1:1000 dilution of an anti-SVN rabbit polyclonal antibody [3]. The amount of bound SVN mutant protein was determined by adding a 1:2000 dilution of donkey anti-rabbit antibody conjugated to horseradish peroxidase. After addition of the horseradish peroxidase substrate buffer and color formation, the reaction was stopped by the addition of 50 µL/well of 2 M H_2SO_4 (after 5 min. for gp120 plate and 15 min. for the gp41 plate) and absorbance was measured at 450 nm for each well.

Additional ELISA assays were performed to test specific monosaccharides and oligomannose 8 (Man-8) for their ability to inhibit SVN-derived mutants from binding to gp120 and gp41. In brief, the plate was prepared as above with glycosylated gp120 or gp41, 100 ng/well of SVN or SNV-derived peptides were added in the presence or absence of increasing concentrations of Man-8 or 100 mM concentrations of the following sugars: glucose, mannose, galactose, xylose, N-acytylglucosamine or 100 µg/well α-acid glycoprotein (to act as a carrier for sialic acid). The plate was then washed and visualized using anti-SVN polyclonal antibodies as above. All data points are averages of triplicate measurements at each concentration.

3. Results

3.1. Purification of recombinant SVN mutants

Since SVN has a disulfide bond linkage between Cys7 and Cys55 which effectively links the first and second domains, the single domain SD1 and SD2 peptides required mutation of Cys7 and Cys55 to another amino acid to avoid possible dimerization and interference with the formation of the remaining disulfide bonds. For this work, both Cys7 and Cys55 were mutated to serine (Fig. 2).

The pET32c(+) vector is designed for cloning and high-level expression of protein sequences fused with the 109aa TRX Tag thioredoxin protein. The thioredoxin expression system was used to enhance disulfide bond formation. A Cloning site containing a cleavable His-Tag sequence was used for easier for detection and purification. The enterokinase digestion site between the SVN peptides and their fusion protein partner provided an easy way to separate SVN mutants from the thioredoxin. Using an SVN purification protocol reported previously [21], purified SVN mutants were obtained and their purities and molecular weights were confirmed by LC–MS, the yield for SVN mutants was also included (Table 1). Although SD1 and SD2 have similar sequences with 75% homology, the SDS-PAGE mobilities of the peptides were different (Fig. 3A, Table 1). SD1 mainly existed in monomeric form, whereas SD2 appears in the form of dimer. The Western blot analysis showed that purified SD1 and SD2 samples were both still recognized by anti-SVN polyclonal antibody (Fig. 3B). Additional dot blot analysis indicated that SD1, SD2 and SVN(1–40) had modestly reduced binding affinity for SVN polyclonal antibodies when compared with SVN (data not shown).

3.2. Disulfide bond pattern of SD1

The disulfide bond pattern of native SVN was previously reported [3]. Recombinant SD1 showed the expected molecular weight of 4944.3 Da by LC–MS. When SD1 was reduced with 10 mM β-mercaptoethanol, its molecular weight increased to 4948.2, a 4 Da increase which demonstrated that SD1 had two disulfide bonds.

Fig. 3 – SDS-PAGE and immunoblot analysis of purified SVN, SD1 and SD2. (A) Coomassie blue-stained gel of the purified peptides. (B) Immunoblot analysis of an identical gel with rabbit anti-scytovirin polyclonal antibodies [Lane 1: SVN, Lane 2: SD1 and Lane 3: SD2].

LC–MS data on the trypsin digest fragments of non-reduced SD1 and SD2 and peptide recognition software [1] were used to identify the disulfide bonds present in SD1 and SD2 (Fig. 4). The molecular weight of one trypsin-digested fragment of recombinant SD1 (m/z = 3154.0 Da) was first input into the software to establish the disulfide bond pattern. This mass matched only the calculated mass for the peptide extending from SD1 residue 1 to SD1 residue 30 bearing an intact disulfide linkage between the only two Cys residues present in that fragment, Cys20–Cys26. The only other possible disulfide bond in SD1 was between Cys32 and Cys38. This bond was confirmed by mass data for the peptide fragment extending from residue 31 to residue 48 (MW = 1790.2Da) which matched that expected for the second disulfide bond. No evidence was seen for different disulfide-bonding patterns between alternate pairs of Cys residues in SD1. This confirmed that the disulfide bond

Fig. 2 – Amino acids sequences of SVN mutants. The cysteines at positions 7 and 55 of the SVN sequence have been mutated into serine.

Table 1 – Production, molecular weight and anti-HIV activities of SVN peptides				
Protein	rEK cleavage efficiency (%)	Yield (mg/L LB medium)	Molecular weight[a]	EC$_{50}$[b] (nM)
SVN	95	10	9712.2	4.5
SD1 (SVN(1–48)Cys7Ser)	98	2.5	4944.3	7.6
SD2 (SVN(49–95)Cys55Ser)	85	1.7	4759.8	182
SVN(1–40)Cys7Ser	96	2.2	4126.4	34.5
SVN(3–45)Cys7Ser	90	1.3	4535.3	>2200
SVN(6–45)Cys7Ser	85	0.8	4281.0	>2200
SVN(11–45)Cys7Ser	89	1.9	3601.1	>2200

[a] Determined by electrospray mass sprectrometry.
[b] Determined against HIV-1$_{RF}$ in a T-lymphocyte cell cytoprotection assay.

arrangement for SD1 was the same as found in native SVN [3]. In a similar manner to SD1, the disulfide bond pattern of recombinant SD2 was established using the trypsin digest fragments corresponding to residues 1–30 (m/z = 3235.8 Da) and residues 31–47 (m/z = 1553.0 Da) of SD2 (Fig. 4).

3.3. Anti-HIV activity of SVN mutants

Native SVN has been reported to display potent antiviral activity against laboratory strains and primary isolates of HIV-1 with EC$_{50}$ values ranging from 0.3 to 22 nM [3]. In side-by-side in vitro anti-HIV assays using CEM-SS host cells and HIV-1$_{RF}$, the recombinant SVN peptide's anti-HIV activities were evaluated and compared to full-length SVN (Table 1). Anti-HIV activity comparable to the full-length SVN was retained in SD1, while SD2 showed about a 40-fold decrease in anti-HIV potency.

Additional experiments examined whether the N- and C-terminus of SD1 were necessary for antiviral activity. A series of deletion mutants were constructed in which 2, 5 or 10 amino acids were deleted from the N-terminus of SD1, or 8 amino acids were deleted from the C-terminus of SD1. Deletion of 2, 5 or 10 N-terminal amino acids completely eliminated antiviral activity, indicating that N-terminal amino acids of SD1 are necessary for maintaining the antiviral activity of SD1. Deletion of the eight C-terminal amino acids resulted in a three- to seven-fold decrease in anti-HIV potentcy, indicating

that the eight C-terminal amino acids are necessary for optimal activity but not as integral as the N-terminal amino acids.

3.4. gp120 and gp41 binding pattern of SVN mutants

We have previously reported that SVN binds to the HIV-1 viral envelope proteins gp120, gp41 and gp160 [3]. In order to compare the binding activity of SVN-derived peptides to SVN for their ability to bind to gp120 and gp41, ELISA experiments were performed. The results showed that SD1 appears to have gp120- and gp41-binding activity similar to intact SVN (Fig. 5). In contrast, SD2 has reduced gp120- (~57%) and gp41- (~44%) binding ability relative to full-length SVN. The binding activity of SVN(1–40) with gp120 and gp41 is about 44 and 70%, respectively, relative to SVN, while the N-terminal mutants have almost no detectable binding activity with gp120 and gp41 (Fig. 6). An additional ELISA experiment showed that 100 mM concentrations of glucose, mannose, galactose, xylose, N-acytylglucosamine and α-acid glycoprotein with sialic acid did not inhibit SVN-derived peptides (SD1, SD2 and SVN(1–40)) binding to gp120 and gp41 (data not shown). Finally, an ELISA experiment demonstrated a concentration-dependent decrease in the binding of SD1 to gp41 as a result of co-incubation with the high-mannose oligosaccharide oligo-mannose-8 (Fig. 7).

SD1(1-48)Cys7Ser:

m/z = 4948.2 Da
2 disulfide bonds   GSGPTYSWNEANNPGGPNRCSNNKQCDGARTCSSSGFCQGTSRKPDPG

m/z = 3154.0 Da
Cys20-Cys26   GSGPTYSWNEANNPGGPNRCSNNKQCDGAR

m/z = 1790.2 Da
Cys32-Cys38   TCSSSGFCQGTSRKPDPG

SD2(49-95)Cys55Ser:

m/z = 3235.8 Da
Cys20-Cys26   PKGPTYSWDEAKNPGGPNRCSNSKQCDGAR

m/z = 1553.0 Da
Cys32-Cys38   TCSSSGFCQGTAGHAAA
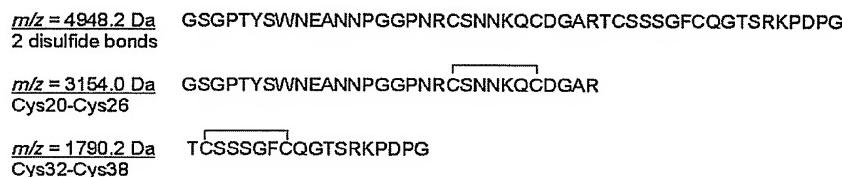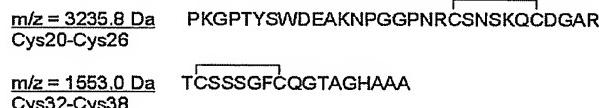
Fig. 4 – Disulfide-bonding pattern of recombinant SD1 and SD2. The values of trypsin-digested SD1 or SD2 were input into the software (http://sx102a.niddk.nih.gov/peptide/) to establish the disulfide-bonding pattern.
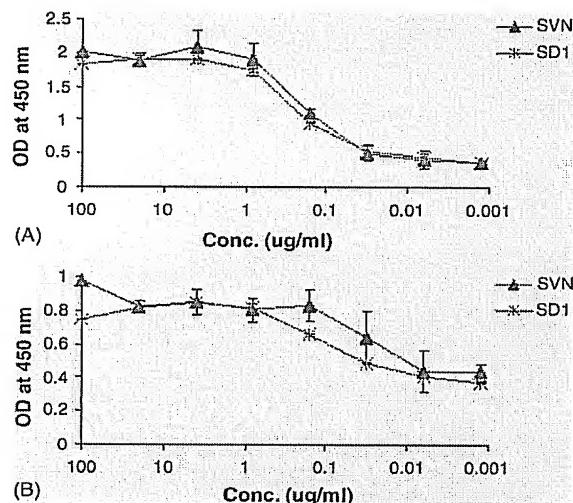
Fig. 5 – ELISA study of the concentration-dependent binding of recombinant SVN or SD1 to (A) gp120 and (B) gp41. Glycosylated gp120 or gp41 was bound to an ELISA plate and then treated with SVN or SD1. The amount of bound SVN or SD1 was determined by absorbance at 450 nm as described in Section 2. Results are the average absorbance at 450 nm (±S.D.) from triplicate wells.

4. Discussion

The anti-HIV protein scytovirin is a novel antiviral protein from a cultured cyanobacterium with little homology to any known protein [3]. SVN does, however, show a strong internal sequence homology leading to the hypothesis that the protein is made up of two putative domains from residue 1–48 (scytovirin domain 1, SD1) and from 49 to 95 (scytovirin domain 2, SD2) (Fig. 1) [3]. Here we tested that hypothesis by cloning and expressing SD1, SD2 and a series of SD1 truncation mutants in an E. coli expression vector. The use of the thioredoxin expression system, necessary for the recombinant production of biologically active SVN [21], was again successful in producing properly folded, biologically active peptides with the appropriate disulfide bonds (Fig. 4). Surprisingly, SD1 displayed equivalent antiviral activity, gp120- and gp41-binding activities relative to SVN (Fig. 5). SD2 however, displayed significantly lower anti-HIV activity than SVN (40-fold higher EC_{50}) and ~50% reduced binding to gp120. Additional truncation of SD1 at the carboxy-terminus by eight amino acids [SVN(1–40)] resulted in a significant decrease in both HIV-1 envelope glycoprotein binding and anti-HIV activity while any additional truncation of SD1 at the N-terminus eliminated biological activity.

The fact that the two domains of SVN differ so significantly in their activities was not expected. Additional structural studies will be crucial for elucidating the exact conformational determinants responsible for the antiviral activity of SVN. Preliminary NMR results have indicated that the SD1 and SD2 domains have some elements of structural similarity (Drs. McFeeters and Byrd, personal communication), but their anti-HIV activities, gp120 and gp41 binding activities are quite
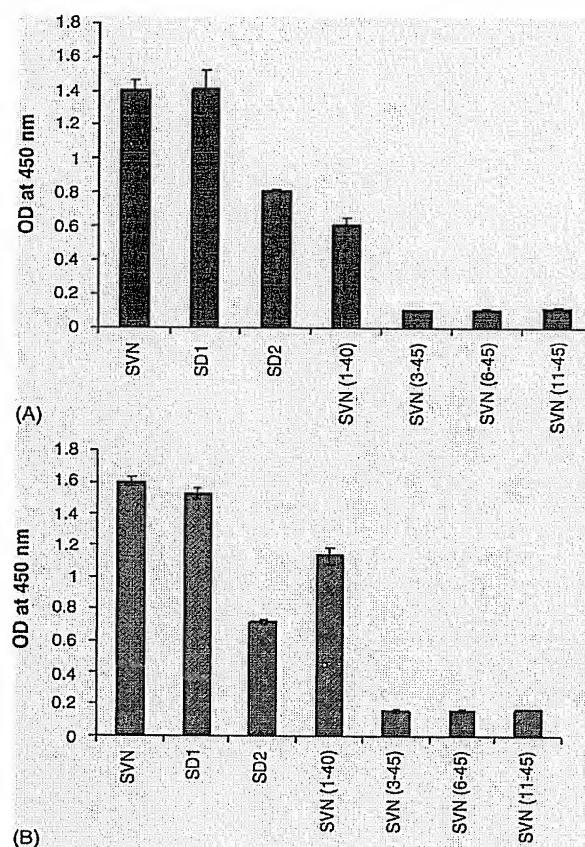
Fig. 6 – ELISA study of recombinant SVN and mutants to (A) gp120 and (B) gp41. Glycosylated gp120 or gp41 was bound to an ELISA plate and then treated with SVN or mutants. The amount of bound SVN or mutants was determined by absorbance at 450 nm as described in the Experimental section. All data are corrected for background antibody absorption in the absence of captured protein (typically < 0.1 OD_{450}). Results are the average absorbance at 450 nm (±S.D.) from triplicate wells.
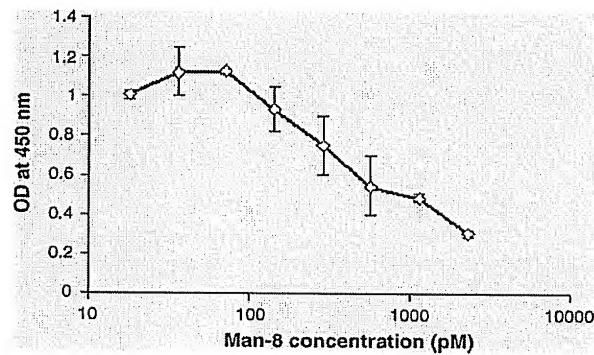
Fig. 7 – Concentration-dependent inhibition of SD1 binding to HIV envelope protein gp41 by the high-mannose oligosaccharide Man-8. Points are the averages ± standard deviations of three replicate determinations.

different. Some insights into these differences can be inferred from the data on the additional truncation mutants and the preliminary NMR data. The information on N-terminal truncations of SD1 and SVN(1–40) indicates that the N-terminal amino acids of SVN are key for the envelope glycoprotein binding and anti-HIV activity of SD1. As the first two amino acids Pro49 and Lys50 in SD2 differ significantly from the Gly1 and Ser2 in SD1 the loss of biological activity following their deletion is non-intuitive and indicated differing structural importance for the amino acids in SD1 vs. those in SD2. Initial data from the NMR studies indicated that the Gly1-Ser2 in SD1 are integral to the structural integrity of domain 1 of SVN and in close proximity to the oligosaccharide binding site while the Pro49-Lys50 in SD2 appear to be part of a hinge region between the two domains of SVN and not of such structural importance (Drs. McFeeters and Byrd, personal communication). The further truncation of SD1 by removal of the eight amino acids at the C-terminus resulted in a significant reduction but not a complete loss of activity. Further deletion of C-terminal amino acids were not performed, since this would disrupt the second disulfide bond of SD1 which likely has a substantial role in defining the conformation and, thus, the biological activity of the peptide. Taken together with the sequence overlap, the data suggest that the C-terminal eight amino acids of SD1 and N-terminal six amino acids of SD2, regions of low sequence homology, may comprise a flexible region separating two oligosaccharide-binding domains of SVN.

SVN has previously been shown to bind to HIV-1 envelope glycoproteins through specific interactions with substructures of high-mannose oligosaccharides [2,3]. In particular, SVN was reported to bind specifically to a α-1-2, α1-2, α1-6-linked tetramannoside [2], a substructure of oligomannose-8 and -9 that are found on the HIV envelope glycoproteins gp120 and gp41. It was therefore important to determine whether or not this level of specificity was replicated in the single domain SVN-derived peptides. To test this, an ELISA experiment was performed in the presence of high concentrations of six different monosaccharides to determine if they could inhibit binding to HIV-1 gp120. None of the sugars was able to inhibit the binding of SD1, SD2 or SVN(1–40) to gp120 (data not shown) thereby indicating that the individual domains retained some level of carbohydrate specificity and were not acting as monosaccharide specific lectins. This result was buttressed by additional experiments showing that oligomannose-8 was able to inhibit the binding of SD1 to gp41 in a concentration-dependent manner (Fig. 7). The hypothesis that SVN specificity resided in the individual domains of SVN was also later confirmed by NMR analysis of a titration of the SVN-binding tetrasaccharide [2] into SVN (Drs. McFeeters and Byrd, personal communication). This level of oligosaccharide specificity in such a relatively small peptide is unique to SD1 among lectins. The molecular interactions necessary to achieve this selective binding are currently being studied.

The development of HIV fusion inhibitors has focused on three strategies: (1) blocking the interaction of gp120 with the cellular receptor CD4, (2) blocking the secondary interaction of gp120 with the cellular co-receptors CCR5 or CXC4, or (3) disrupting the formation of the six-helix bundle of fusion-active gp41 [19,20]. T-20, a member of the latter group was has been approved by the U.S. Food and Drug Administration for treatment of infection by HIV-1 [6,13,14]. Antibodies binding to gp120 such as Pro 542 (a fusion protein including 4 IgG2 molecules in which the variable fragments of both light chains are substituted with the D1/D2 domains of CD4) [11] and peptides such as CD4 M33 (a 27-amino acid peptide mimicking the CD4 domain D1) [15] are also in clinical trials. The peptide SD1 with essentially the same anti-HIV and gp120 binding activities as full-length SVN represents a distinct strategy for an anti-HIV therapeutic or prophylactic agent. The advantages of the SD1 peptide are significant. SD1 has 3 fewer disulfide bonds than SVN and also potentially less immunogenicity due to its decreased size. Furthermore, SD1 is within the size range of peptides easily synthesized using automated techniques, which provides an alternative to biological production. These advantages, in addition to its broad spectrum of activity against HIV-1 viruses, low nanomolar activity and physical stability, provide ample reason for the continued study of SD1 for its potential utility as an anti-HIV microbicide or therapeutic.

REFERENCES

[1] http://sx102a.niddk.nih.gov/peptide.
[2] Adams EW, Ratner DM, Bokesch HR, McMahon JB, O'Keefe BR, Seeberger PH. Oligosaccharide and glycoprotein microarrays as tools in HIV glycobiology; glycan-dependent gp120/protein interactions. Chem Biol 2004;11(6):875–81.
[3] Bokesch HR, O'Keefe BR, McKee TC, Pannell LK, Patterson GM, Gardella RS, et al. A potent novel anti-HIV protein from the cultured cyanobacterium Scytonema varium. Biochemistry 2003;42(9):2578–84.
[4] Botos I, Wlodawer A. Proteins that bind high-mannose sugars of the HIV envelope. Prog Biophys Mol Biol 2005;88(2):233–82.
[5] Boyd MR, Gustafson KR, McMahon JB, Shoemaker RH, O'Keefe BR, Mori T, et al. Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential application to microbicide development. Antimicrob Agents Chemother 1997;41(7):1521–30.
[6] Chen RY, Kilby JM, Saag MS, Enfuvirtide. Expert Opin Investig Drugs 2002;11:1837–43.
[7] Cole AM. Antimicrobial peptide microbicides targeting HIV. Protein Pept Lett 2005;12(1):41–7.
[8] Gordon YJ, G EG, McDermott AM. A review of antimicrobial peptides and their therapeutic potential as anti-infective drugs. Curr Eye Res 2005;30(7):505–15.
[9] Gulakowski RJ, McMahon JB, Staley PG, Moran RA, Boyd MR. A semiautomated multiparameter assay for anti-HIV drug screening. J Virol Methods 1991;33:87–100.
[10] Han Z, Xiong C, Mori T, Boyd MR. Discovery of a stable dimeric mutant of cyanovirin-N (CV-N) from a T7

phage-displayed CV-N mutant library. Biochem Biophys Res Commun 2002;292(4):1036–43.

[11] Jacobson JM, Israel RJ, Lowy I, Ostrow NA, Vassilatos LS, Barish M, et al. Treatment of advanced human immunodeficiency virus type 1 disease with the viral entry inhibitor PRO 542. Antimicrob Agents Chemother 2004;48:423–9.

[12] Kilby JM, Hopkins S, Venetta TM, DiMassimo B, Cloud GA, Lee JY, et al. Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. Nat Med 1998;4(11):1302–7.

[13] Lalezari JP, Eron JJ, Carlson M, Cohen C, DeJesus E, Arduino RC, et al. A phase II clinical study of the long-term safety and antiviral activity of enfuvirtide-based antiretroviral therapy. AIDS 2003;17(5):691–8.

[14] Lalezari JP, Henry K, O'Hearn M, Montaner JS, Piliero PJ, Trottier B, et al., TORO 1 Study Group. Enfuvirtide, an HIV-1 fusion inhibitor, for drug-resistant HIV infection in North and South America. N Engl J Med 2003;348(22):2175–85.

[15] Martin L, Stricher F, Missé D, Sironi F, Pugnière M, Barthe P, et al. Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. Nat Biotechnol 2003;21:71–6.

[16] Mori T, Shoemaker RH, Gulakowski RJ, Krepps BL, McMahon JB, Gustafson KR, et al. Analysis of sequence requirements for biological activity of cyanovirin-N, a potent HIV (human immunodeficiency virus)-inactivating protein. Biochem Biophys Res Commun 1997;238(1):218–22.

[17] O'Keefe BR. Biologically active proteins from natural product extracts. J Nat Prod 2001;64(10):1373–81.

[18] O'Keefe BR, Shenoy SR, Xie D, Zhang W, Muschik JM, Currens MJ, et al. Analysis of the interaction between the HIV-inactivating protein cyanovirin-N and soluble forms of the envelope glycoproteins gp120 and gp41. Mol Pharmacol 2000;58:982–92.

[19] Ryser HJ, Fluckiger R. Progress in targeting HIV-1 entry. Drug Discov Today 2005;10(16):1085–94.

[20] Tamamura H, Otaka A, Fujii N. Development of anti-HIV agents targeting dynamic supramolecular mechanism: entry and fusion inhibitors based on CXCR4/CCR5 antagonists and gp41-C34-remodeling peptides. Curr HIV Res 2005;3(4):289–301.

[21] Xiong C, O'keefe BR, Botos I, Wlodawer A, McMahon JB. Overexpression and purification of scytovirin, a potent, novel anti-HIV protein from the cultured cyanobacterium *Scytonema varium*. Protein Expr Purif 2006;46(2):233–9.

JMB

ELSEVIER

The Novel Fold of Scytovirin Reveals a New Twist For Antiviral Entry Inhibitors

Robert L. McFeeters[1], Changyun Xiong[2], Barry R. O'Keefe[2]
Heidi R. Bokesch[2,3], James B. McMahon[2], Daniel M. Ratner[4]
Riccardo Castelli[5], Peter H. Seeberger[5] and R. Andrew Byrd[1]*

[1]*Structural Biophysics Laboratory, National Cancer Institute, Frederick MD 21702-1201, USA*

[2]*Molecular Targets Development Program National Cancer Institute Frederick, MD 21702-1201 USA*

[3]*SAIC-Frederick, National Cancer Institute, Frederick MD 21702-1201, USA*

[4]*Department of Chemistry Massachusetts Institute of Technology, Cambridge MA 02139, USA*

[5]*Laboratory for Organic Chemistry, Swiss Federal Institute of Technology (ETH) Zürich, 8092 Zürich Switzerland*

Corresponding author

The solution structure of the potent 95 residue anti-HIV protein scytovirin has been determined and two carbohydrate-binding sites have been identified. This unique protein, containing five structurally important disulfide bonds, demonstrates a novel fold with no elements of extended regular secondary structure. Scytovirin contains two 39 residue sequence repeats, differing in only three amino acid residues, and each repeat has primary sequence similarity to chitin binding proteins. Both sequence repeats form similarly structured domains, with the exception of one region. The result is two carbohydrate-binding sites with substantially different affinities. The unusual fold clusters aromatic residues in both sites, suggesting a binding mechanism similar to other known hevein-like carbohydrate-binding proteins but differing in carbohydrate specificity. Scytovirin, originally isolated from the cyanobacterium *Scytonema varium*, holds potential as an HIV entry inhibitor for both therapeutic and prophylactic anti-HIV applications. The high-resolution structural studies reported are an important initial step in unlocking the therapeutic potential of scytovirin.

© 2007 Elsevier Ltd. All rights reserved.

Introduction

Protein:carbohydrate interactions play fundamental roles throughout biology. Bacterial infection, cell growth, inflammation, cell mobility, fertilization,

cell to cell adhesion, and viral infection are all influenced by protein:carbohydrate interactions. Therefore protein:carbohydrate interactions are a topic of major interest throughout science. One well-studied group of lectins is the hevein-like family. This family is structurally united by the presence of similar chitin-binding domains (CBDs) and are members of carbohydrate-binding module family 18.[1] Hevein, isolated from the latex of *Hevea brasiliensis*, is composed of 43 residues and its structure has been known since 1991.[2] Other hevein-like CBDs vary in length, typically from 38–45 residues, and are often found in multiple repeats within a single protein. A dual CBD-like motif is present in the potent antiviral protein scytovirin, originally isolated from the cyanobacterium *Scytonema varium*.

Present address: D.M. Ratner, Department of Molecular and Cell Biology, Boston Medical Center, Boston University, Boston, MA, 02215, USA.

Abbreviations used: CBD, chitin-binding domain; RDC, residual dipolar coupling; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; HSQC, heteronuclear single quantum coherence.

E-mail address of the corresponding author: rabyrd@ncifcrf.gov

Scytovirin is of particular interest due to the fact that it possesses potent and broad spectrum antiviral, including anti-HIV, activity not found in the other hevein-like domain proteins.[3] Scytovirin exhibits low nanomolar activity against HIV and binds to the HIV envelope glycoproteins gp160, gp120 and gp41, but it does not bind to the T-cell extracellular CD4 receptor or other common cell surface proteins.[4] The interaction is carbohydrate-dependent, with scytovirin most readily binding to glycosylated gp120. Scytovirin's ability to completely inactivate laboratory strains and primary isolates of HIV-1[4] makes it a promising candidate for anti-HIV microbicide development.

Scytovirin contains two 39 residue repeats, which exhibit rather low similarity to a subset of diverse CBD-containing proteins.[4] The initial sequence analysis offers little indication of the mechanism of antiviral activity beyond similarity to CBDs. The comparison to hevein-like CBDs was not originally examined; however, the sequence similarity between the scytovirin repeats and the hevein domain is approximately 30%. In addition, several features common to hevein family members[5,6] are present in scytovirin: There are two CBD-like domains separated by a proline-rich linker. Both scytovirin and hevein-like CBDs have a high percentage of cysteine residues, all participating in disulfide bonds. Each domain of scytovirin has three aromatic residues, similar to the conserved triad of aromatic residues found in other hevein-like CBDs. Thus, it seems plausible that scytovirin is another hevein family member.

However, scytovirin shows distinct differences from the hevein-like proteins. First, scytovirin does not bind chitin.[4] In fact, scytovirin shows a very restrictive carbohydrate binding specificity, having no measurable affinity for monosaccharides or common trisaccharide cores from higher order mannose oligosaccharides[4,7] (*vide infra*). This observation is in stark contrast to the increasing affinity for longer oligosaccharides reported for other hevein-like CBDs.[8] Second, the disulfide bonds among the conserved cysteine residues of scytovirin are shuffled compared to the consensus hevein arrangement. Scytovirin has only five cysteine residues per domain. These align exactly with five of the eight cysteine residues per domain in the classical hevein arrangement. The presence of an odd number of cysteine residues in each domain of scytovirin causes the disulfide bonds to rearrange, resulting in two disulfides per domain and one inter-domain disulfide,[4] unlike the four disulfide bonds typically found in hevein domains. Although scytovirin has three aromatic residues per domain, similar to hevein-like CBDs, the sequence arrangement of the aromatic residues does not match the classic hevein domain sequential arrangement, *vide infra*.

Thus, scytovirin has remained a rather enigmatic, but potent antiviral protein. In order to better understand the antiviral mechanism and enable development of this natural protein, we have determined the three-dimensional solution structure of recombinant scytovirin and found it to have a novel fold, quite different from hevein. We have also examined the carbohydrate-binding properties with short oligomannose carbohydrates, which were previously identified.[7] These structural studies are a vital part of the future development for scytovirin and detail the unique characteristics of this protein, both with respect to its antiviral activity and its relationship to other well-studied carbohydrate-binding proteins.

Results

Chemical shift and NOE assignments

Chemical shift assignments of recombinant scytovirin (Figure 1(a)) were complicated by resonance degeneracy. A doubling of most peaks was immediately apparent from the ^{15}N-heteronuclear single quantum coherence (HSQC) (Figure 1(b)). Duplicity of the amide backbone chemical shifts for a majority of the resonances readily suggested that the sequence repeats of scytovirin have similar backbone folds. There was an indication that each sequence repeat might represent a separate domain of the total protein structure, and we refer to the two sequence repeats as domains, which is borne out in the complete analyses (*vide infra*). Chemical shift degeneracy was exacerbated for side-chains where ^{13}C and ^{1}H resonances were not as well dispersed as ^{15}N chemical shifts. For illustration, strip plots for a stretch of five consecutive residues in each domain, showing the C^{α} and C^{β} resonances and sequential correlations from triple resonance NMR spectra, are shown in Figure 1(c).

The three amino acid differences between sequence repeats of scytovirin aided the sequential assignment process; nevertheless, meticulous manual analysis of standard triple resonance spectra had to be supplemented with additional experiments to unambiguously assign backbone resonances. An HMQC-NOESY-HMQC,[9] twice taking advantage of ^{15}N chemical shift dispersion, greatly aided the resolution of inter- *versus* intra-domain backbone sequential assignments. Pulse sequences to assign through proline resonances[10] were also used to confirm assignments. Information from these experiments was especially useful for the P-G-G-P sequence in each domain and the P-D-P-G-P sequence in the linker. Manual analysis of nuclear Overhauser effect spectroscopy (NOESY) data was also utilized to help resolve chemical shift ambiguity. In summary, nearly complete domain-specific resonance assignments were obtained for the first 94 residues of scytovirin. Residue 95, the C-terminal alanine, was not observable. Over 98% of backbone and 92% of side-chain resonances were accounted for. As expected, each ^{15}N-HSQC peak pair corresponded to complementary resonances of the 39 residue sequence repeats. Chemical shift
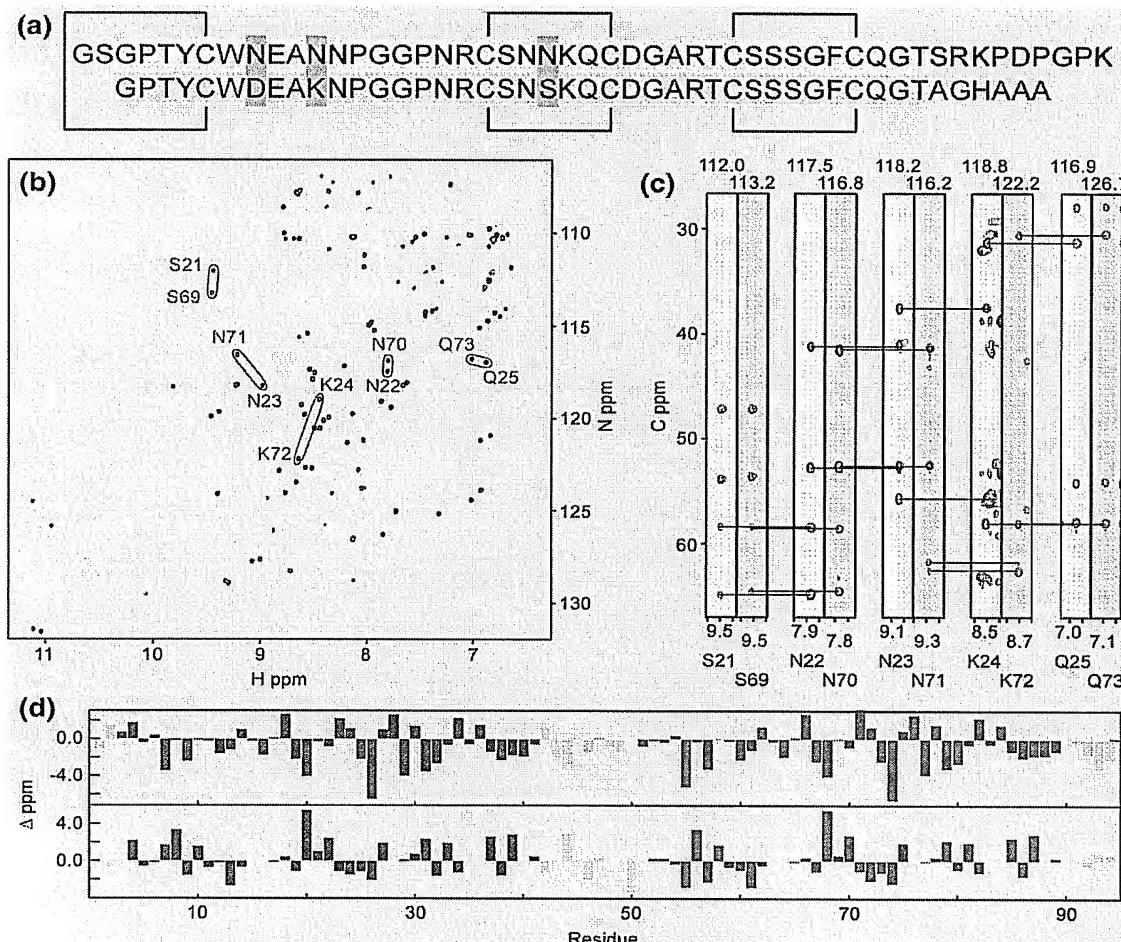
Figure 1. Scytovirin sequence, resonance assignments, and chemical shift indices. (a) Amino acid sequence of scytovirin aligned to show sequence repeats. The background of the first domain (SD1) is shaded in red while the background of the second domain (SD2) is shaded in blue. Black lines indicate disulfide bonds. Green boxes highlight the three natural sequence differences between domains. (b) An ^{15}N-HSQC of scytovirin illustrates peak doubling of complementary resonances. Residues S21, N22, N23, K24, and Q25 from SD1 are labeled in red and complementary residues S69, N70, S71, K72, and Q73 from SD2 are labeled in blue. (c) Slices from HNCACB and CBCA(CO)NH spectra for the same ten residues as (b). HNCACB C^α resonances are shown in red and C^β resonances in blue. CBCA(CO)NH resonances, from previous residue C^α and C^β resonances, are shown in magenta. Background shading is red for SD1 and blue for SD2. (d) The difference in C^α (top) and C^β chemical shifts from random coil values. No extended regions of regular secondary structure are expected, since the characteristic positive difference for α-helices and negative difference for β-sheets is not observed for C^α chemical shifts.

analysis of backbone C^α and C^β resonances[11,12] (Figure 1(d)) gave the first indication that no extended regions of regular secondary structure were present.

NOESY assignments were also complicated by chemical shift degeneracy, especially for inter- *versus* intra-domain NOEs. No symmetry for the two domains was assumed and an initial structure was calculated from a sparse set of unambiguously assigned NOEs. A moderate resolution structure was obtained and a degree of 2-fold symmetry was immediately apparent. The initial structure allowed for resolution of more inter- *versus* intra-domain NOE ambiguity. After multiple iterations, 1205

intraresidue, 1006 sequential, and 634 long range NOEs were assigned (Table 1).

Structural overview and backbone flexibility

The solution structure of scytovirin was determined using NOEs and residual dipolar coupling data[13] (RDCs). Initial structures, determined using only NOEs, exhibited a backbone RMSD of 1.4(±0.1) Å to the mean structure. The sequence repeats corresponded to structured domains with residues 3–43 forming structural domain 1 (SD1) and 51–89 forming structural domain 2 (SD2). The disulfide bonding pattern, previously determined from pro-

Table 1. NMR and refinement statistics for scytovirin

Distance constraints	
Total NOE	2905
Intra-residue	1247
Inter-residue	1658
Sequential ($|i-j| = 1$)	679
Medium-range ($|i-j| < 4$)	324
Long-range ($|i-j| > 5$)	655
Total RDCs	148
NH	82
$H^\alpha C^\alpha$	66
Q (%)	0.25
Structure statistics	
Violations (mean and s.d.)	
Distance constraints > 0.5 Å (Å)	24±2
Average violation > 0.5 Å (Å)	0.69±0.16
Max. distance constraint violation (Å)	1.17
Deviations from idealized geometry	
Bond lengths (Å)	0.01±0.06
Bond angles (°)	1.59±2.55
Impropers (°)	0.08±2.19
Average r.m.s.d. to mean[a] (Å)	
Heavy	0.42±0.11
Backbone	0.51±0.11

[a] Calculated from 20 refined structures.

teolysis and peptide mapping,[5] was readily apparent in these initial structures. Structure calculations with and without disulfide restraints showed only small differences, and subsequent calculations included these restraints. Backbone RMSDs for the domains, aligned individually, were 1.0 Å for SD1 and 1.3 Å for SD2. More striking than the domain similarity was the complete absence of regular secondary structure. In agreement with chemical shift indexing data, no extended α-helices or β-sheets were observed. Circular dichroism (Supplementary Data Figure 1) also indicates an atypical fold. The CD spectrum for scytovirin does not exhibit an α-helical, β-sheet, mixed helix/sheet, or unfolded profile. Refinement against *J*-modulated $H^\alpha C^\alpha$ and NH RDCs confirmed the unusual fold and total lack of regular secondary structure. Two regions of extended structure are superficially similar to β-sheets. In SD1, residues 28–32 lie antiparallel to residues 38–42. Also, residues 6–12 lie antiparallel to the disjoint sections 17–20 and 26–28. Although reminiscent of β-sheets, neither Phi/Psi angles nor chemical shift indexing strictly indicate the presence of classical β-sheet folds. Furthermore, temperature coefficients of the amide resonances[14] are not consistent with hydrogen bonding patterns expected for antiparallel β-sheets. The same holds for complementary residues in SD2. Additionally, there are no Ile, Leu, or Val residues in the sequence, which results in a very minimal hydrophobic core. The limited hydrophobic core within each domain is consistent with the unusual fold. Thus, scytovirin has no recognizable α-helix or β-sheet elements of secondary structure.

The lowest energy 20 refined scytovirin structures are shown in Figure 2 and coordinates are deposited under PDB accession number 2JMV. In the final

structures, the backbone RMSD of the entire protein improved to 0.42(±0.11) Å (Table 1). Backbone RMSDs of the individual domains improved to 0.25 Å for SD1 and 0.40 Å for SD2. In addition to improved precision, RDC refinement helped define the linker and last loop-like structure of each domain. Calculated *versus* observed RDC plots and cross-validation are included as Supplementary Data Figure 2. A DALI[15] structure search found no similar structures for either SD1, SD2 or full-length scytovirin. Thus scytovirin represents a truly novel fold.

Despite the lack of regular secondary structure, scytovirin appears to be quite ordered on NMR measurable timescales. Transverse and longitudinal ^{15}N relaxation rates showed no regions of significant variation except in the C terminus. The correlation time derived from these data was 5.1 ns, which corresponds to a monomeric 9.7 kDa protein. Heteronuclear NOEs[16] at multiple field strengths resulted in average values (excluding the two C-terminal residues) of 0.80±0.05 and 0.80±0.07 at 600 and 800 MHz, respectively. Data are included as Supplementary Figure 3. No significant change in average NOE is observed when calculated separately for the individual domains. It is apparent that the backbone of scytovirin does not exhibit flexibility, except the last two C-terminal residues. The N terminus is well ordered, beginning at residue 1, and participates in structuring of SD1. The interactions of residues G1 and S2 with C26, D27, and R30 contribute to the structural difference between domains. Hydrogen/deuterium exchange experiments show that 17 residues in each domain exhibit protected amide protons. All of the protected amides were on the interior of the protein. Some of
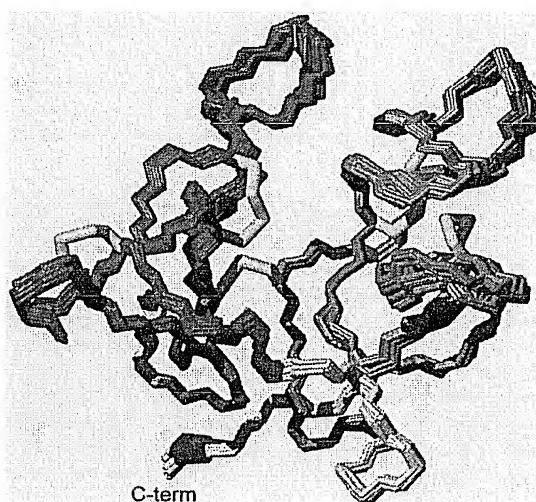
Figure 2. Solution structure of scytovirin. Backbone traces of lowest energy 20 scytovirin structures of 200 calculated from refinement against $H^\alpha C^\alpha$ and NH RDCs. SD1 is colored red, SD2 blue, disulfide bonds in yellow, and the linker and termini gray.

these amides must participate in hydrogen bonds; however, due to the lack of regular secondary structure it was not possible to unambiguously assign hydrogen bonding partners. Therefore, the structures were calculated without such constraints.

Domain interface

The interface between the SD1 and SD2 lies in the middle of the protein and helps stabilize the fold of each domain. The structure of each domain was calculated separately, using intra-domain NOEs. The backbone RMSDs worsened due to the loss of inter-domain NOEs, but the same overall folds were recognized. One end of the interface is cross-linked by the disulfide bond between C7 and C55. Multiple unambiguous NOEs, most notable T5 HG to T53 HG, substantiate the presence of the disulfide bond. The remainder of the interface is predominantly hydrophobic. A large number of unambiguous NOEs are observed between residue C20 and residue Q73. The interface is ~580 Å^2 and may help stabilize the domain structures, in agreement with initial studies of the individual domains (unpublished results). The linker does not appear to contribute significantly to the interface.

Conserved structural elements

As expected from the similarity of chemical shifts in each domain, the amino acid sequence repeats have similar overall backbone folds; however, the RMSD between SD1 and SD2 is 5 Å. Close inspection reveals a conformational change that explains the difference between the domains. Before discussion of the differences, it is beneficial to understand the likenesses.

Three discreet, common structural elements are apparent in both domains of scytovirin (Figure 3(a)). The first conserved structural element, the "top loop," is formed by residues 12–17(60–65) of SD1 (SD2) (Figure 3(b)). The top loop circumnavigates approximately 240° of a complete circle. The first position of the top loop is the site of one of the three sequence differences (N12/K60) between SD1 and SD2, and the rest of the loop is formed by a P-G-G-P sequence. Both top loops exhibit a similar pattern of heteronuclear NOEs at 600 and 800 MHz. The first glycine (G15/G63) in the loop has an average NOE value of 0.77±0.01, whereas the second glycine (G16/G64) has a value of 0.84±0.02. The first glycine is closer to the solvent-exposed apex of the loop and therefore it is reasonable that it is slightly less conformationally restricted. The second conserved structural element, denoted the "middle turn," is formed by residues 22–25(70–73) (Figure 3(c)). A disulfide bond between cysteine residues 20(68) and 26(74) helps demark the boundaries of this conserved structural element. In SD1, a probable hydrogen bond between the backbone carbonyl oxygen of residue 22 and the backbone amide proton of residue 25 produces a likeness to a classical turn. However, unlike a classical turn, the last residue of both middle turns introduces an abrupt change in direction. The change in direction is followed by a
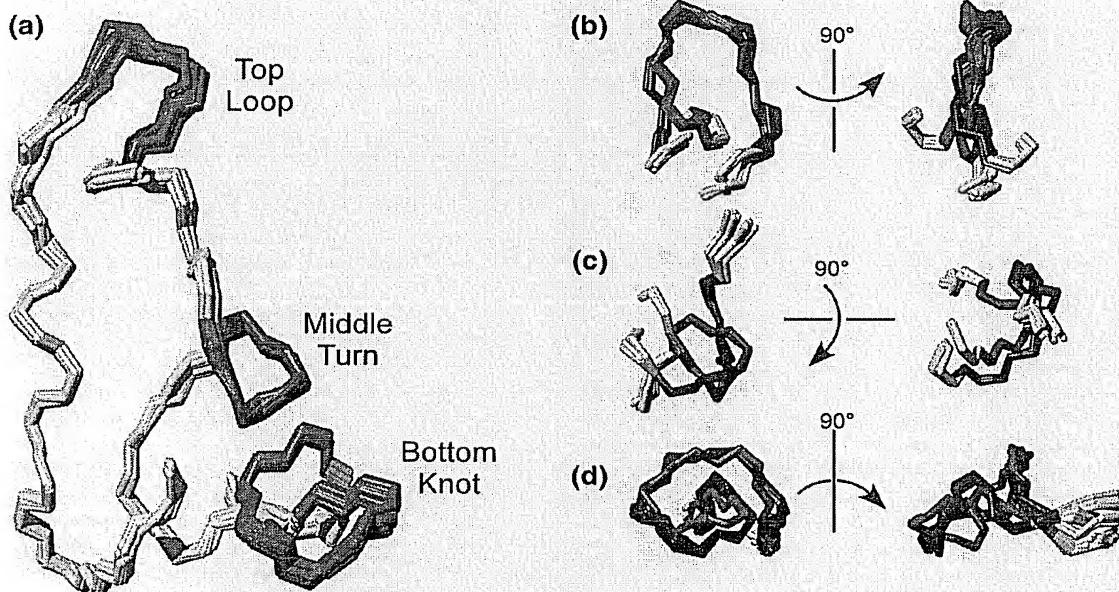
Figure 3. Three major sub-structural elements in scytovirin. (a) Twenty member ensemble of SD1 colored to show the three major loop-like elements conserved in both domains in red. See the text for discussion. (b) The top loop, (c) middle turn and (d) bottom knot of SD1 (red) and SD2 (blue) are aligned to illustrate the structural similarities and differences.

kinked extended region leading to the third conserved structural element and most distinguishing feature of scytovirin, denoted the "bottom knot," which is formed from residues 31–39(79–87) (Figure 3(d)). As with the middle turn, a disulfide bond between residues 32–38 (80–86) helps define the boundaries of this conserved structural element. Refinement against RDCs greatly improves resolution of this element in each domain, revealing a quite unusual fold.

Domain differences

Both SD1 and SD2 are composed of the same structural elements; however, the relative orientations of the elements within each domain are not the same, resulting in the 5 Å RMSD between SD1 and SD2. The major differences are centered around the middle turn. In SD2, the middle turn is flipped 180° with respect to the rest of the domain when compared to the middle turn of SD1 (Figure 3(c)). Backbone chemical shift differences between SD1 and SD2 are greatest for residues in the middle turn and the extended region directly following (Figure 4). Interactions between residues G1 and C26, D27, R30 of scytovirin are responsible for the difference, since the side-chains intimately pack in this region of SD1 and help determine the orientation of the turn. The same does not hold for SD2, since residues P49 and K50 in the linker, equivalent in position to G1 and S2 for SD1, pack and interact differently with the residues near the middle turn of SD2. Residues preceding P49 in the linker, which are absent at the N terminus, may also contribute to the differences. The final result is a flipped orientation of the middle turn. The structural significance of the two N-terminal residues, G1 and S2, is supported by the decrease in

anti-HIV EC_{50} by greater than a factor of 500 when these two residues are removed.[17]

As a consequence of the reversed middle turn, the top loop is also reversed between domains. The right-handedness of SD1 is mirrored by the left-handedness of SD2. Changes in side-chain packing caused by the natural residue differences of N9 to D57 before the top loop and N12 to K60 at the beginning of the top loop contribute to the reversed handedness. Differences also occur in hydrophobic contributions at position A11/A59. For SD1, the domain interface provides some protection for the A11. For A59 of SD2, protection by the interface is not available so the hydrophobic side-chain buries in the domain, positioning the residue much differently. These relatively small effects combine to cause the regions between the top loop and middle turn to be different between SD1 and SD2. These conformations are confirmed by the observed RDCs for these regions, and the difference leads to the large backbone RMSD between the two domains.

The backbone folds of SD1 and SD2 are quite similar, except for the region between the top loop and middle turn. The concerted change in handedness of the top loop and 180° reversal of the middle turn effectively cancel each other, preserving the remainder of the overall fold. The backbone RMSD of the two domains is 2.5 Å, when the top loop through the middle turn are left out of the calculation. The major structural differences are thereby relegated to only part of the protein, and do not dramatically impact residues near the carbohydrate-binding site (*vide infra*). This allows scytovirin to position residues critical for carbohydrate binding similarly between domains. Thus, both domains of scytovirin retain the ability to bind oligosaccharide, albeit with different affinities.

Oligosaccharide binding

Scytovirin has been shown to bind specific oligosaccharides found on gp41 and gp120.[4] In particular, scytovirin was shown to bind to a specific tetrasaccharide substructure of the high mannose oligosaccharide normally found decorating HIV-1 envelope glycoproteins.[7] The Manα(1→2)Manα (1→6)Manα(1→6)Man tetrasaccharide is pictured in Figure 5(a). Chemical shift perturbation mapping of scytovirin was used to monitor titration of this tetrasaccharide (see Figure 5(b)). Different exchange kinetics were observed for residues in SD1 *versus* residues in SD2. At 500 MHz, residues in SD1 exhibit intermediate exchange, whereas residues in SD2 exhibit fast exchange. These exchange regimes can be shifted by conducting the experiments at 800 MHz, where the intermediate exchange becomes slow and the fast exchange becomes intermediate (data not shown). The results suggest significantly different affinities for binding of the tetrasaccharide to each domain. Chemical shifts as a function of ligand concentration were fit to a two-site model (Matlab
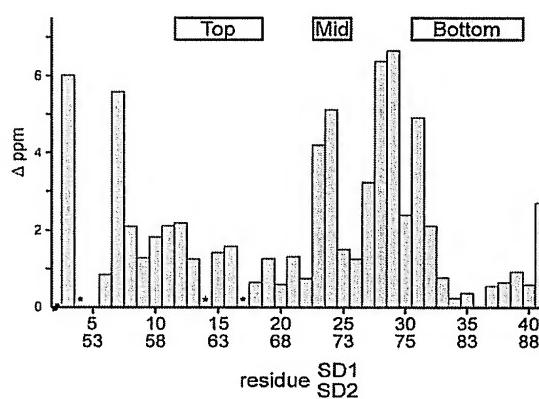
Figure 4. Amide backbone chemical shift difference between domains. The absolute difference of amide backbone chemical shifts for equivalent residues in SD1 and SD2 are plotted. ^{1}H contributions have been scaled to ^{15}N ppm. Asterisks mark the position of proline residues. At the top, boxes demark residues composing the top loop, middle turn, and bottom knot.

(a)

(b)

(c)
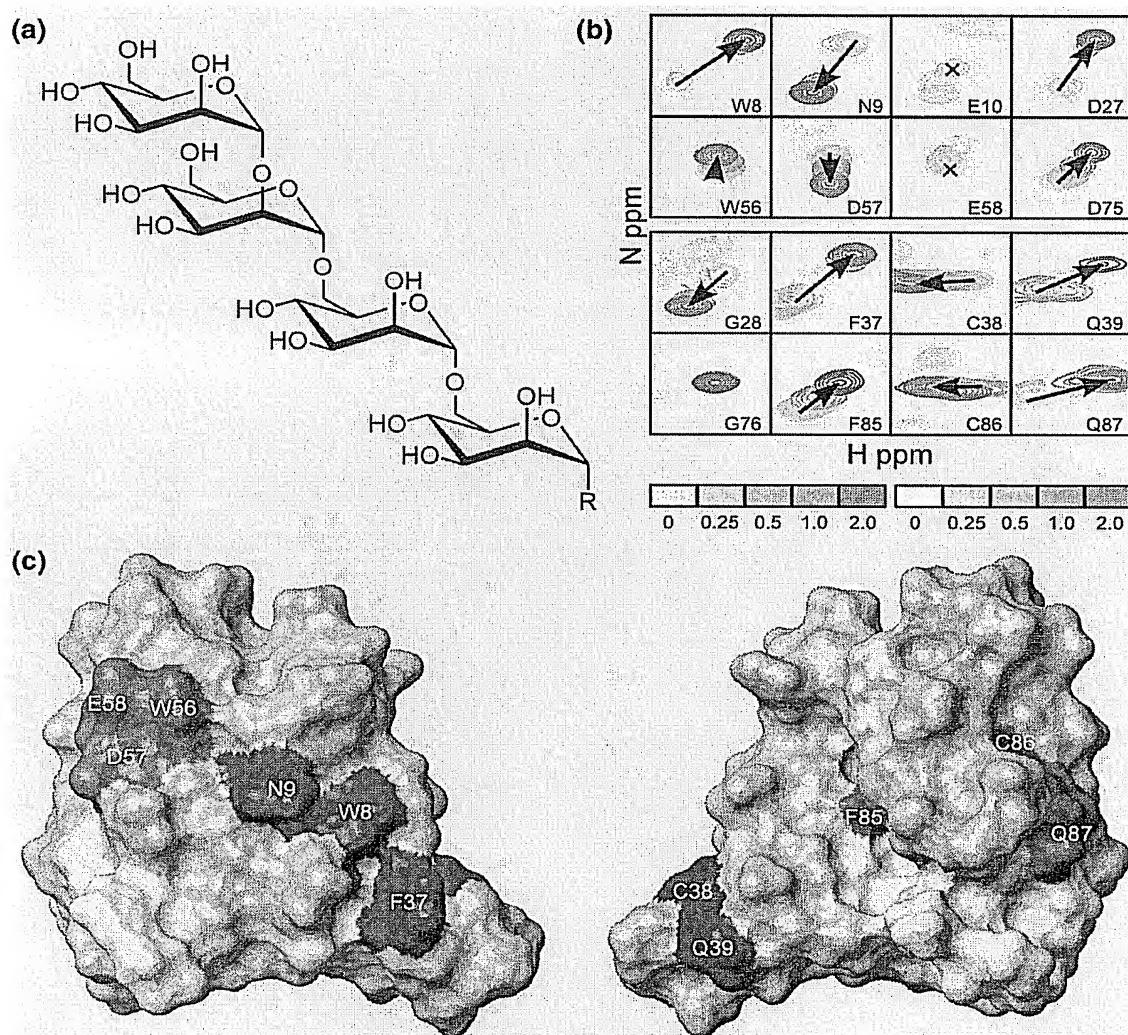
Figure 5. Oligosaccharide titrations reveal binding site. (a) Structure of the tetrasaccharide used for NMR binding studies, where R is the same as in Adams *et al.*[7] (b) Overlay of sections from ^{15}N-HSQC spectra collected with increasing amounts of tetrasaccharide. Complementary residues showing the largest chemical shift perturbation are shown. Residues from SD1 are in red and residues from SD2 are in blue. Color gradients from light to dark correspond to tetrasaccharide:scytovirin ratios of 0:1, 0.25:0,0.5:1,1:1,2:1. (c) Surface representation of scytovirin (SD1 pink, SD2 light blue) colored to show residues having largest change in chemical shift upon tetrasaccharide binding (SD1 red, SD2 blue).

algorithm kindly provided by Dr David Fushman, University of Maryland) providing dissociation constant estimates of 30 μM for SD1 and 160 μM for SD2 (Supplementary Data Figure 4). These measurements agree with data measured from isothermal titration calorimetry (S. Shenoy & B. R. O'Keefe, personal communication).

Residues that demonstrate backbone amide chemical shift changes upon addition of tetrasaccharide are almost identical for SD1 and SD2. Mapping those residues onto the scytovirin structure is shown in Figure 5(c). It appears that each domain of scytovirin binds oligosaccharide in a similar fashion. This agrees with biochemical studies of the individually expressed domains.[17] There are three aromatic residues in each domain (SD1, Y6, W8, F37; and SD2, Y54, W56, F85), and, with the exception of F85 in SD2, all the aromatic residues of scytovirin are clustered near the binding sites (Figure 5(c)). Both $N^{\varepsilon 1}H^{\varepsilon 1}$ of tryptophan W8(SD1) and W56(SD2) demonstrate chemical shift perturbations, thus suggesting involvement in tetrasaccharide binding. The clustering of aromatics agrees well with an existing model in which two aromatic residues, separated by one residue in the amino acid sequence, play a critical role in one class of protein–oligosaccharide interactions.[18]

Oligosaccharide-binding site comparison

The differences between the two oligosaccharide-binding sites can partially be explained from the apo structure of scytovirin. The N9 residue in SD1, which is the site of side-chain substitution for D57 in SD2, shows large backbone amide chemical shift changes in the presence of tetrasaccharide, implicating involvement in binding. The negative charge of D57 may contribute to decreased oligosaccharide affinity in SD2. Also, the residues between the top loop and middle turn show minor chemical shift perturbations upon addition of tetrasaccharide, and some structural differences between the two domains may contribute to differences in binding. The proximity of G1 and S2 to surface-exposed residues with the largest chemical shift perturbations in SD1 indicate the significance of the ordered N terminus and local structure that contribute to the different oligosaccharide-binding affinity. The lack of chemical shift perturbation for G76 in SD2 differs from the significant change observed for the equivalent G28 in SD1. Thus, even though the sites are quite similar, enough difference exists to cause an appreciable difference in oligosaccharide affinity. Furthermore, both sites exhibit considerable specificity. We have confirmed the previous observation that a related $Man\alpha(1{\rightarrow}6)Man\alpha(1{\rightarrow}6)Man$ trisaccharide[7] does not bind to scytovirin based on the absence of observed chemical shift perturbations in NMR titration experiments.

Discussion

At first glance, scytovirin appears to be another member of the hevein-like family, since it binds oligosaccharides, has a majority of the signature CBD cysteine residues[2] (Figure 6(a)), and has a similar aromatic triad involved in carbohydrate binding. The likeness of scytovirin to CBDs, however, does not extend to the backbone fold. The scytovirin structure is quite distinct from that of the hevein domain (Figure 6(b)). The difference in structure is, in part, due to the different disulfide bond formation. The absence of a cysteine at position 17 in scytovirin, strictly conserved in all hevein family members, causes a dramatic rearrangement of disulfide bonding (see Figure 6(a)). In scytovirin, the disulfide bonding pattern is C20–C26 and C32–C38 in SD1, C68–C74 and C80–C86 in SD2 with an inter-domain C7–C55 disulfide bond.[4] This corresponds to a sequential disulfide bonding pattern between the 2nd–3rd and 4th–5th (7th–8th, 9th–10th in SD2) cysteine residues with one inter-domain disulfide bond, between the 1st–6th cysteine residues. The disulfide bonds are very important to the structural integrity of the two domains, and the inter-domain disulfide bond assists in bringing the two domains together to form the overall globular structure. In all hevein family members, the pattern of disulfide bonds is not sequential, but intertwined as evidenced from disulfide bonding between the

conserved 1st–4th, 2nd–5th, 3rd–6th, and 7th–8th cysteine residues (Figure 6(a)). The conserved disulfide pairing makes it reasonable to assume that the hevein domain structure is conserved across all family members, as represented by the *B. arreguin* hevein domain (Figure 6(b)). The hevein domain exhibits short α-helical and β-strand secondary structures, contrary to the absence of such structural elements in scytovirin.

The sequence repeat, or dual hevein motif, of scytovirin is also seen in other CBDs. For example, a recently described tomato lectin shows the same duplication of hevein domains joined by proline-rich linkers.[5] The tomato lectin has no inter-domain disulfide bond, hence there is no *a priori* expectation that the two domains would contact or interact with one another. The dual motif would be anticipated to have a fold comprised of two typical hevein domains connected by a linker, as suggested by the model in Figure 6(b). Conversely, the inter-domain disulfide bond of scytovirin is reminiscent of mistletoe lectin where two CBDs are held together by a lone, inter-domain disulfide bond.[6] The conserved hevein disulfide pairing of the tomato lectin (Figure 6(a)) would suggest that the individual domains have the hevein fold and that they are simply tethered. It is not known if there is any contact or interaction between the two domains, as seen in scytovirin. This variation in domain structure and interconnection might provide further differentiation in carbohydrate recognition specificity; however, further investigations are required to examine this possibility.

Each domain of scytovirin contains a triad of aromatic residues that is similar to the triad in hevein domains and is known to be responsible for carbohydrate binding in hevein.[18] Despite the sequence similarity to CBDs, scytovirin lacks measurable binding affinity for chitin, monosaccharides, or common trisaccharides.[4,7] Scytovirin does not bind the $GlcNAc\beta(1{\rightarrow}4)GlcNAc$ structure of chitin but instead shows high specificity for a tetrasaccharide structure of $Man\alpha(1{\rightarrow}2)Man\alpha(1{\rightarrow}6)Man\alpha$ $(1{\rightarrow}6)Man$. One plausible explanation for these differences is the altered arrangement of key aromatic residues in the binding sites of scytovirin compared to hevein domains. Interestingly, SD1 of scytovirin preserves the structural arrangement of the aromatic triad (Y6, W8, and F37) and has a nearly identical affinity for its tetrasaccharide-binding partner as hevein does for $(GlcNAc)_4$.[8] SD2 exhibits a lower apparent binding affinity for the tetrasaccharide, and the aromatic triad (Y54, W56, and F85) is not as tightly clustered. The side-chain of F85 is not readily able to participate in binding due to the reversed arrangement of the middle turn. Thus, we predict that scytovirin and hevein utilize similar protein:carbohydrate interactions; however a different specificity is achieved as a direct result of the altered disulfide bonding pattern and rearrangement of the aromatic triads in each domain. In binding experiments with the tetrasaccharide, the two binding sites are independent with

(a)

```
SVN SD1    GPTYCWNEANNPGGPN- -RCSNNKQCDG-ARTCSSSGFCQGT---

SVN SD2 ---GPTYCWDEAKNPGGPN- -RCSNSKQCDG-ARTCSSSGFCQGT


HEV        EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKD
TL-CBMn1   PRCGAQGDGGNCPSGMCCSVWGWCGKTYGYCAP-QNCQKQCPA---
TL-CBMn2 ---GRCGWQADGKSCPNGKCCSYGGWCGTTSDYCAR-QNCQKQCIL
```

(b)
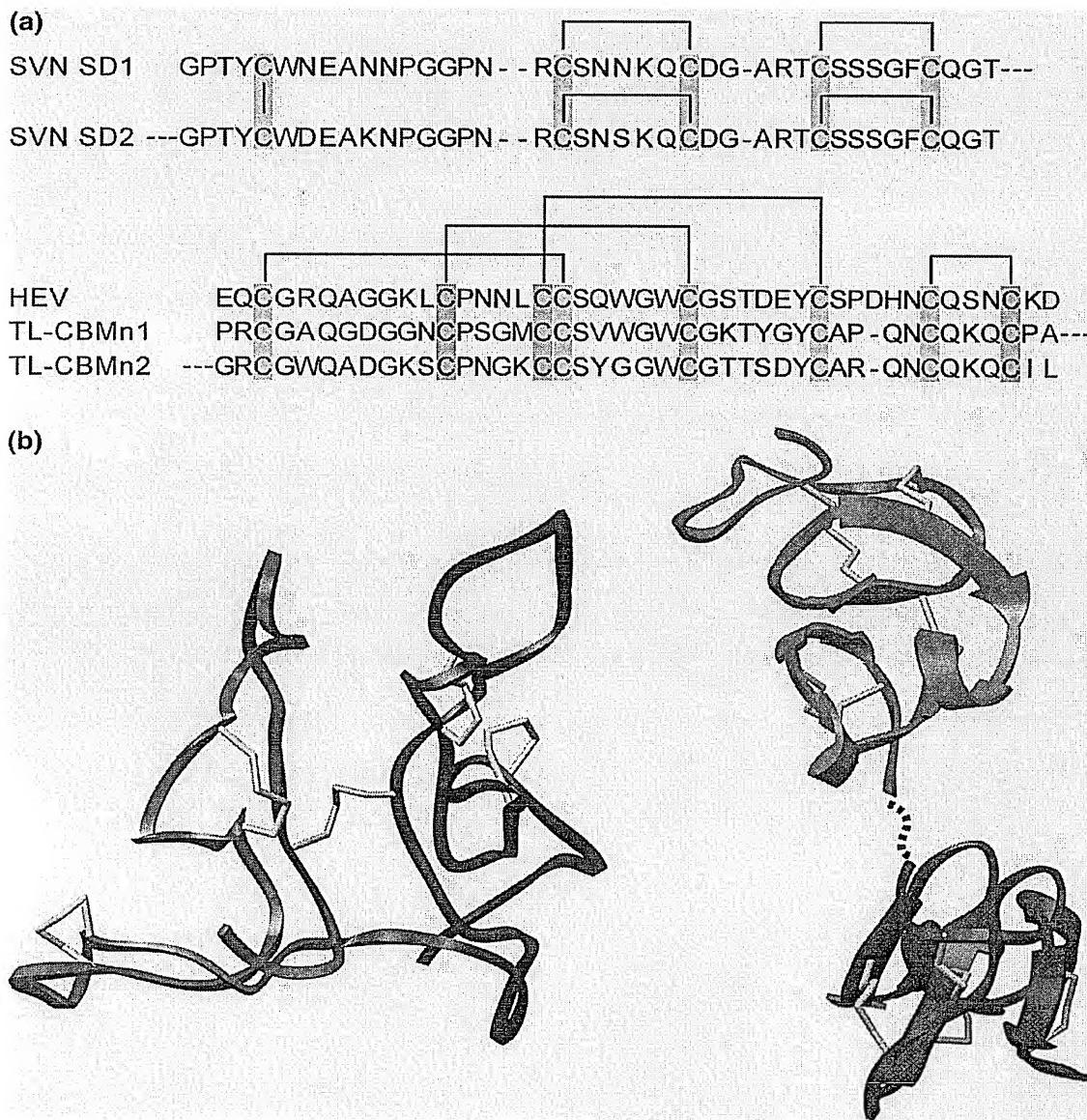
Figure 6. Sequence alignment, disulfide bonding pattern, and domain comparison of scytovirin and hevein-domains. (a) An amino acid sequence alignment of both scytovirin domains (SVN SD1 and SD2), hevein (HEV), and the hevein-like tomato lectin N-terminal chitin-binding modules[5] (TL-CBMn1 and TL-CBMn2) illustrates the drastically different disulfide bond formation of scytovirin. Disulfide bonds are indicated by lines. Gold boxes indicate cysteine residues conserved in scytovirin and hevein-like proteins, whereas blue boxes indicate cysteine residues only found in hevein-like CBDs. (b) Comparison of the backbone folds of the dual CBDs of scytovirin (left) to a model of two arbitrarily oriented hevein domains (PDB ID: 1HEV) (right). All cysteine residues side-chains are shown, colored in yellow. SVN SD1 is shown in red and SVN SD2 is shown in blue. The model of a dual hevein-domain protein without any inter-domain disulfide bond (right) is formed from two copies of 1HEV and the linker is represented as a broken line.

relatively weak affinity. As noted for other CBDs,[8] increased affinity is achieved by binding to large carbohydrates or glycoproteins that can deliver carbohydrates to both scytovirin domains, and this multivalency is likely responsible for the high anti-HIV activity of scytovirin.

The presence of two weak carbohydrate-binding sites that result in potent anti-viral activity towards gp41 and gp120 parallels the observations for another cyanobacterial lectin protein with potent antiviral activity, cyanovirin-N.[19] Cyanovirin-N exhibits two binding sites with micromolar affinity, similar to scytovirin. The cyanovirin-N-binding sites, however, do not involve aromatic residues, based on both NMR titrations[20] and the crystal structure of the cyanovirin-N:Man-9 complex.[21]

Similar to scytovirin, cyanovirin-N exhibits very high affinity toward viral glycoprotein epitopes. A recent study of cyanovirin-N[19] observed precipitation of gp41 by cyanovirin-N that could be prevented by mutation of one of the binding sites, suggesting that the precipitation was due to cross-linking but concluding that the nanomolar potency was probably not due to cross-linking. Instead, the multivalency may be related to the spatial specificity of glycoprotein epitopes. Our data suggest that scytovirin may act in a parallel fashion, but the nature of the carbohydrate-binding sites are different and lead to a distinct carbohydrate specificity compared to cyanovirin-N.[7]

The structure of scytovirin introduces a novel fold and new twist to carbohydrate-binding proteins. Knowledge of the carbohydrate-binding sites provides important opportunities for structural engineering to improve this molecule as an antiviral, and particularly anti-HIV agent. Further studies are in progress to better understand the protein:carbohydrate complex. A greater understanding of how scytovirin binds to viral glycoprotein carbohydrates will contribute to advanced development of antiviral therapies and potential topical microbicide prophylaxis.

Materials and Methods

Sample preparation and NMR data acquisition

Scytovirin was expressed and purified as described.[22] The NMR buffer used for all experiments was 20 mM Mes (pH 5.5), 100 μM EDTA, 10 μM NaN$_3$. The concentration of scytovirin was 1 mM for the unbound sample and 0.5 mM for the sample used for carbohydrate titration. All data, unless otherwise noted, were collected at 25 °C. The titration, assignment, and ^{13}C-edited NOESY data were collected on Varian Inova spectrometers with cryogenically cooled probes at field strengths of 500, 600 and 800 MHz, respectively. ^{15}N-edited NOESY and proline sequential assignment data were collected at 600 MHz on a Bruker Avance spectrometer. ^{15}N-edited and ^{13}C-edited NOESY mixing times were 100 and 125 ms, respectively. Heteronuclear ^{15}N-^{1}H NOEs were collected with a recycle delay of 5.5 s or recycle delay of 2.5 s and a saturation time of 3 s. Hydrogen/deuterium exchange spectra were collected at times of 20 min, 40 min, 1.5 h, 4 h, 16 h, and four days after a lyophilized sample of scytovirin was resuspended in ^{2}H$_2$O. All data were processed using nmrPipe[23] and visualized using SPARKY.[24]

Structure calculation and refinement against RDCs

Structure calculations were conducted using XPLOR-NIH.[25] Due to the lack of extended secondary structure predicted from C$^\alpha$/C$^\beta$ chemical shifts, database calculated dihedral angle restraints were not used. A total of 1206 intraresidue, 1005 sequential, and 634 long range (defined as NOEs between protons of residues five or greater positions apart in the sequence) were used in the calculation. The lowest energy 20 structures of 200 calculated are reported. The precise and consistent suite of J-

modulated pulse sequences[13] were used to measure NH and H$^\alpha$C$^\alpha$ RDCs. Integer multiples of 0.70 and 1.0 ms were used for H$^\alpha$C$^\alpha$ and NH coupling delays, respectively. Sixty-seven H$^\alpha$C$^\alpha$ and 82 NH couplings were used for structural refinement. Ramachandran statistics included 13 dihedral angles in the most favorable region, 32 in the additionally allowed, 19 generously allowed and five disallowed.

Oligosaccharide titration

Oligosaccharides were synthesized as reported.[26,27] Titration data were collected at both 500 and 800 MHz in NMR buffer. Ratios of approximately 0:1, 0.5:1, 1:1 and 2:1 were collected for the Manα(1-2)Manα(1-6)Manα(1-6) Man tetrasaccharide. Concentrations of 0:1, 0.5:1, 1:1, 2:1 and 10:1 were collected for the non-binding Manα(1-6) Manα(1-6)Man trisaccharide. At 800 MHz, data were collected at temperatures of 10, 17.5 and 25 °C for temperature dependence of oligosaccharide binding and tentative indication of hydrogen bonding.

Protein Data Bank accession code

Coordinates have been deposited in the RCSB Protein Data Bank under accession number 2JMV.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2007.03.030

References

1. Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). Carbohydrate-binding modules: fine-

tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781.

2. Rodriguez-Romero, A., Ravichandran, K. G. & Soriano-Garcia, M. (1991). Crystal structure of hevein at 2.8 A resolution. *FEBS Letters*, **291**, 307–309.

3. Charan, R. D., Munro, M. H., O'Keefe, B. R., Sowder, R., McKee, T. C., Currens, M. J. et al. (2000). Isolation and characterization of *Myrianthus holstii* lectin, a potent HIV-1 inhibitory protein from the plant *Myrianthus holstii*(1). *J. Nat. Products*, **63**, 1170–1174.

4. Bokesch, H. R., O'keefe, B. R., McKee, T. C., Pannell, L. K., Patterson, G. M., Gardella, R. S. et al. (2003). A potent novel anti-HIV protein from the cultured cyanobacterium *Scytonema varium*. *Biochemistry*, **42**, 2578–2584.

5. Peumans, W. J., Rouge, P. & Van Damme, E. J. (2003). The tomato lectin consists of two homologous chitin-binding modules separated by an extensin-like linker. *Biochem. J.* **376**, 717–724.

6. Stoeva, S., Franz, M., Wacker, R., Krauspenhaar, R., Guthohrlein, E., Mikhailov, A. et al. (2001). Primary structure, isoforms, and molecular modeling of a chitin-binding mistletoe lectin. *Arch. Biochem. Biophys.* **392**, 23–31.

7. Adams, E. W., Ratner, D. M., Bokesch, H. R., McMahon, J. B., O'Keefe, B. R. & Seeberger, P. H. (2004). Oligosaccharide and glycoprotein microarrays as tools in HIV glycobiology; glycan-dependent gp120/protein interactions. [see comment]. *Chem. Biol.* **11**, 875–881.

8. Asensio, J. L., Canada, F. J., Siebert, H. C., Laynez, J., Poveda, A., Nieto, P. M. et al. (2000). Structural basis for chitin recognition by defense proteins: GlcNAc residues are bound in a multivalent fashion by extended binding sites in hevein domains. *Chem. Biol.* **7**, 529–543.

9. Ikura, M., Bax, A., Clore, G. M. & Gronenborn, A. M. (1990). Detection of nuclear Overhauser effects between degenerate amid proton resonances by heteronuclear three-dimensional NMR spectroscopy. *J. Am. Chem. Soc.* **112**, 9020–9022.

10. Schubert, M., Ball, L. J., Oschkinat, H. & Schmieder, P. (2000). Bridging the gap: a set of selective ^{1}H-^{15}N-correlations to link sequential neighbors of prolines. *J. Biomol. NMR*, **17**, 331–335.

11. Wishart, D. S. & Sykes, B. D. (1994). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR*, **4**, 171–180.

12. Wishart, D. S., Sykes, B. D. & Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, **31**, 1647–1651.

13. McFeeters, R. L., Fowler, C. A., Gaponenko, V. V. & Byrd, R. A. (2005). Efficient and precise measurement of H(alpha)-C(alpha), C(alpha)-C', C(alpha)-C(beta) and H(N)-N residual dipolar couplings from 2D H (N)-N correlation spectra. *J. Biomol. NMR*, **31**, 35–47.

14. Cierpicki, T., Zhukov, I., Byrd, R. A. & Otlewski, J. (2002). Hydrogen bonds in human ubiquitin reflected in temperature coefficients of amide protons. *J. Magn. Reson.* **157**, 178–180.

15. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.

16. Kay, L. E., Torchia, D. A. & Bax, A. (1991). Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, **28**, 8972–8979.

17. Xiong, C., O'Keefe, B. R., Byrd, R. A. & McMahon, J. B. (2006). Potent anti-HIV activity of scytovirin domain 1 peptide. *Peptides*, **27**, 1668–1675.

18. Chavez, M. I., Andreu, C., Vidal, P., Aboitiz, N., Freire, F., Canada, F. J. & Jimeriez-Barbero, J. (2005). On the importance of carbohydrate-aromatic interaction for the molecular recognition of oligosaccharides by proteins: NMR studies of the structure and binding affinity of AcAMP2-like peptides with non-natural naphtyyl and fluoroaromatic residues. *Chem. Eur. J.* **11**, 7060–7074.

19. Barrientos, L. G., Matei, E., Lasala, F., Delgado, R. & Gronenborn, A. M. (2006). Dissecting carbohydrate-Cyanovirin-N binding by structure-guided mutagenesis: functional implications for viral entry inhibition. *Protein Eng. Des. Select.* **19**, 525–535.

20. Bewley, C. A. & Otero-Quintero, S. (2001). The potent anti-HIV protein cyanovirin-N contains two novel carbohydrate binding sites that selectively bind to Man(8) D1D3 and Man(9) with nanomolar affinity: implications for binding to the HIV envelope protein gp120. *J. Am. Chem. Soc.* **123**, 3892–3902.

21. Botos, I., O'Keefe, B. R., Shenoy, S. R., Cartner, L. K., Ratner, D. M., Seeberger, P. H. et al. (2002). Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J. Biol. Chem.* **277**, 34336–34342.

22. Xiong, C., O'keefe, B. R., Botos, I., Wlodawer, A. & McMahon, J. B. (2006). Overexpression and purification of scytovirin, a potent, novel anti-HIV protein from the cultured cyanobacterium *Scytonema varium*. *Protein Expr. Purif.* **46**, 233–239.

23. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. [see comment]. *J. Biomol. NMR*, **6**, 277–293.

24. Goddard, T. D. & Kneller, D. G. (0000). *SPARKY 3*. University of California, San Francisco.

25. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73.

26. Plante, O. J., Buchwald, S. L. & Seeberger, P. H. (2000). Halobenzyl ethers as protecting groups for organic synthesis. *J. Am. Chem. Soc.* **122**, 7148–7149.

27. Ratner, D. M., Plante, O. J. & Seeberger, P. H. (2002). A linear synthesis of branched high-mannose oligosaccharides from the HIV-1 viral surface envelope glycoprotein gp120. *Eur. J. Org. Chem.* **5**, 826–833.

Structures of the Complexes of a Potent Anti-HIV Protein Cyanovirin-N and High Mannose Oligosaccharides*

Istvan Botos‡, Barry R. O'Keefe§, Shilpa R. Shenoy§, Laura K. Cartner¶, Daniel M. Ratner‖**,
Peter H. Seeberger‖‡‡, Michael R. Boyd§§, and Alexander Wlodawer‡¶¶

*From the ‡Macromolecular Crystallography Laboratory, NCI, National Institutes of Health, Frederick,
Maryland 21702-1201, §Molecular Targets Drug Discovery Program, Center for Cancer Research, NCI–Frederick,
National Institutes of Health, Frederick, Maryland 21702, ¶Intramural Research Support Program, SAIC-Frederick,
Frederick, Maryland 21702, ‖Department of Chemistry, Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139, and the §§USA Cancer Research Institute, College of Medicine, University of South Alabama,
Mobile, Alabama 36688*

The development of anti-human immunodeficiency virus (HIV) microbicides for either topical or *ex vivo* use is of considerable interest, mainly due to the difficulties in creating a vaccine that would be active against multiple clades of HIV. Cyanovirin-N (CV-N), an 11-kDa protein from the cyanobacterium (blue-green algae) *Nostoc ellipsosporum* with potent virucidal activity, was identified in the search for such antiviral agents. The binding of CV-N to the heavily glycosylated HIV envelope protein gp120 is carbohydrate-dependent. Since previous CV-N-dimannose structures could not fully explain CV-N-oligomannose binding, we determined the crystal structures of recombinant CV-N complexed to Man-9 and a synthetic hexamannoside, at 2.5- and 2.4-Å resolution, respectively. CV-N is a three-dimensional domain-swapped dimer in the crystal structures with two primary sites near the hinge region and two secondary sites on the opposite ends of the dimer. The binding interface is constituted of three stacked α1→2-linked mannose rings for Man-9 and two stacked mannose rings for hexamannoside with the rest of the saccharide molecules pointing to the solution. These structures show unequivocally the binding geometry of high mannose sugars to CV-N, permitting a better understanding of carbohydrate binding to this potential new lead for the design of drugs against AIDS.

Of the more than 30 million people infected with HIV[1] before 1997, 75–85% acquired the virus through heterosexual con-

tacts (1); thus AIDS is likely to continue to affect the general population. The development of a vaccine active against multiple clades of HIV is complicated by the high mutation rate of the virus (2). In the absence of vaccines, there is a growing interest in the development of anti-HIV virucides for either topical or *ex vivo* use (3). A unique natural product identified in the search for new antiviral agents was cyanovirin-N (CV-N), originally isolated from cultures of the cyanobacterium (blue-green algae) *Nostoc ellipsosporum* (4).

Nanomolar concentrations of CV-N potently inactivate diverse strains of HIV-1, HIV-2, SIV, and FIV (4, 5), acting at the level of the virus, not the target cell, to abort the infection process. This is achieved by preventing essential interactions between the envelope glycoprotein and target cell receptors (4–6). For HIV-1, glycoprotein gp120 is primarily involved in cell entry with approximately half of its molecular weight provided by carbohydrates (7). Out of the 24 *N*-linked oligosaccharides found on its surface, 11 are high mannose or hybrid type (8). Studies have shown that the binding of CV-N to gp120 is carbohydrate-dependent (4, 6, 9). Moreover, CV-N also binds free *N*-linked oligosaccharides, having nanomolar affinity for the D1D3 isomer of Man-8 and oligomannose-9 (Man-9) (10–12) and directly competing with gp120 for CV-N binding. NMR and isothermal titration calorimetry experiments have shown that the binding sites exhibit different affinities for the high mannose oligosaccharides (13).

CV-N, a 101-amino acid protein, exists in solution as either a compact monomer or a dimer, whereas all crystal structures show it exclusively as three-dimensional domain-swapped dimers (14–19). Three-dimensional domain swapping is an oligomerization process in which two or more protein chains exchange identical domains (20).

Monomeric CV-N consists of two similar domains with an overall ellipsoidal shape. Although the sequence of the protein is duplicated with over 60% identity between residues 1–50 and 51–101, domain definitions based on the primary and tertiary structures do not coincide exactly. Sequence-defined domain A consists of residues 1–50, whereas residues 51–101 form domain B. Each domain contains mostly β-strands and loops. Two intramolecular disulfide bonds (C8-C22, C58-C73) are important for the structural stability and anti-HIV activity (21). A change of torsion angles in the hinge region (residues 49–54) separates domains A and B of CV-N into an extended form in

[1] The abbreviations used are: HIV, SIV, and FIV are human, simian, and feline immunodeficiency viruses, respectively; CV-N, cyanovirin-N; gp120, 120-kDa surface envelope glycoprotein of HIV; gp41, 41-kDa transmembrane subunit of HIV envelope; ITC, isothermal titration

calorimetry; CHES, 2-(cyclohexylamino)ethanesulfonic acid; CAPS, 3-(cyclohexylamino)-1-propanesulfonic acid; CAPSO, 3-(cyclohexylamino)-2-hydroxypropanesulfonic acid.
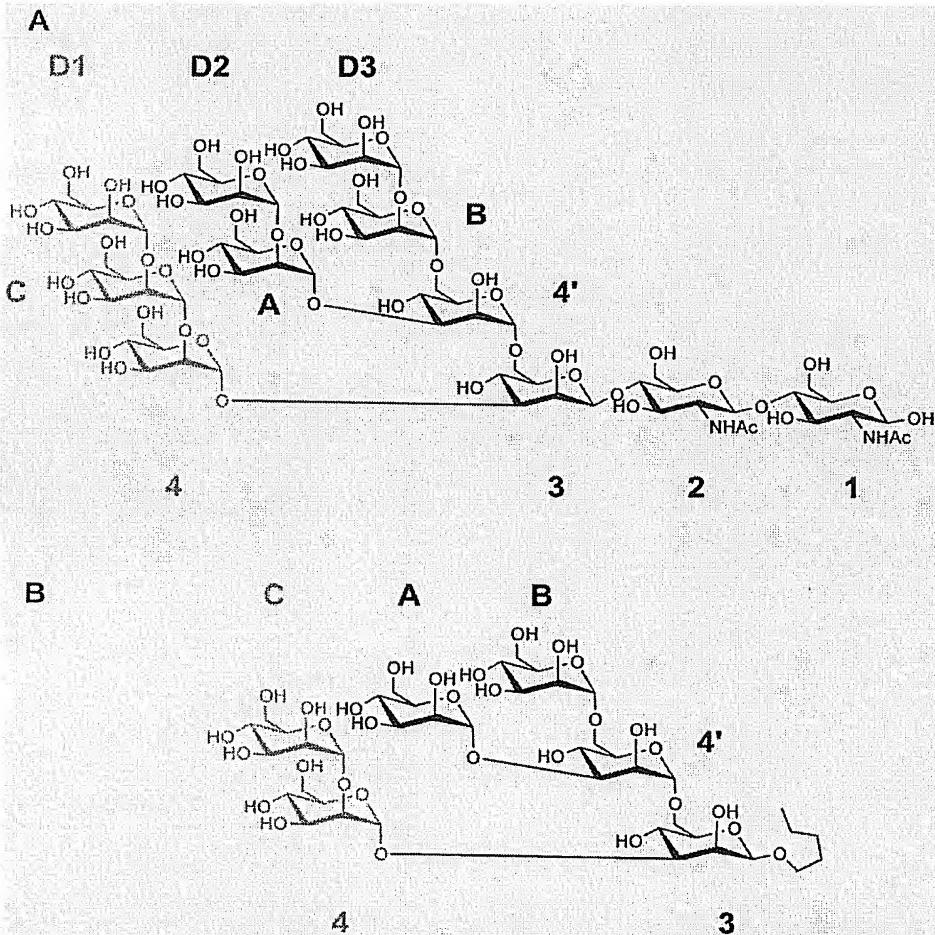
FIG. 1. **Chemical structures of oligosaccharides.** Man-9 (*A*) and synthetic hexamannoside (*B*), used in these studies, are displayed with standard *numbering* and the binding interface highlighted in *green*.

which they do not contact each other. A domain-swapped dimer is formed by two such symmetrically related extended monomers in which domain A comes in contact with domain B′ from a different chain, the overall structure of each pseudo-monomer (AB′ and A′B) being virtually identical to the compact monomer with the exception of the hinge residues. Although the structures of individual pseudo-monomers resemble the compact monomer very closely, the relative orientation of the pseudo-monomers vary widely between different structures (19, 22). Crystal packing seems to thermodynamically favor the dimeric form, selectively trapping this form in the growing crystals, decreasing the abundance of the dimer in solution, and shifting the equilibrium from monomer to dimer.

Since the antiviral activity of CV-N results from its binding to carbohydrates, understanding the structural basis of such interactions is important for the potential development of this protein as an anti-AIDS agent. In the present study, we report medium resolution structures of the complexes of wild-type CV-N with the oligosaccharide Man-9 (Fig. 1A) and with a synthetic hexamannoside (Fig. 1B).

EXPERIMENTAL PROCEDURES

High Mannose Saccharides—Man-9 was purchased from Glyko Inc. (Novato, CA). Solution phase synthesis of the branched hexamannoside has been described previously (23).

Isothermal Titration Calorimetry Experiments—The calorimetric titrations were performed on a VP™-isothermal titration calorimetry (ITC) titration calorimeter from Microcal, Inc. (Northampton, MA). In a typical experiment, 10-μl aliquots of a CHES solution were injected from a 250-μl syringe into a rapidly mixing (300 rpm) solution of CV-N (cell volume, 1.3472 ml). Control experiments involved injecting identical amounts of the CHES solution into buffer without CV-N. The concentration of CV-N was 0.463 mM, and that of CHES was 6.48 mM. Titrations were carried out at 30 °C in 50 mM sodium phosphate buffer, 0.2 M NaCl, 0.02% NaN$_3$, pH 7.5. The isotherms, corrected for dilution/buffer effects, were fit using the Origin ITC Analysis software according to the manufacturer's protocols. A nonlinear least square method was used to fit the titration data and to calculate the errors. From the binding curve values for enthalpy, stoichiometry, and binding affinity were extracted. Thermodynamic parameters were calculated using $\Delta G = -RT \ln K_a$, $\Delta G = \Delta H - T\Delta S$.

Protein Purification and Crystallization—Wild-type CV-N was cloned, expressed, and purified from *Escherichia coli* as reported previously (24). Only crystals grown at high pH in silica hydrogel (Hampton Research) were successfully utilized for preparation of complexes with oligosaccharides. These crystals were grown from droplets containing an equal mixture of protein and precipitant (1 M sodium citrate and 0.1 M CHES, pH 10.3). After 14 days at room temperature, the crystals grew to 0.2 × 0.15 × 0.1 mm. Crystals were soaked for 48 h in 1 mM Man-9 or hexamannose immediately before data collection. CV-N-hexamannose co-crystals grew under a range of conditions under which the wild-type CV-N alone would not crystallize.

Crystallographic Procedures—X-ray data for the CV-N-oligosaccha-

TABLE I
Crystallographic statistics

	Man-9	Hexamannoside
Diffraction data statistics		
Cell parameters	a = b = 61.51, c = 147.94	a = b = 61.35, c = 147.56
	$\alpha = \beta = \gamma = 90.0$	$\alpha = \beta = \gamma = 90.0$
Space group	$P4_12_12$	$P4_12_12$
Molec/asymmetric unit	1	1
Resolution (Å)	2.5	2.4
Total reflections	214,251	129,116
Unique reflections	7,967	11,728
Completeness (%)	99.9	99.8
Last shell	100.0	99.6
Avg. I/σ	23.0	12.1
R_{merge} (%)[a]	7.1	5.6
Refinement statistics		
R-factor (%)[b]	24.2	22.4
R_{free} (%)[c]	29.8	28.9
R.m.s.d. bonds (Å)[d]	0.009	0.01
Angles (deg)	1.6	1.8

[a] $R_{merge} = \Sigma|I - \langle I \rangle|/\Sigma I$, where I is the observed intensity, and $\langle I \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections after rejections.
[b] R-factor $= \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.
[c] R_{free} defined in Ref. 38.
[d] R.m.s.d., root mean square deviation.

ride complexes were collected at 100 K on beamline X9B, National Synchrotron Light Source (NSLS), Brookhaven National Laboratory, with the ADSC Quantum4 CCD detector, at 0.92-Å wavelength. A cryoprotectant, consisting of 80% mother liquor and 20% ethylene glycol, was used. Data were processed with HKL2000 (25) (Table I). Molecular replacement was carried out with AmoRe (26) using a compact monomer as a search model. With the correct solution, the symmetry-related mates were generated, and although domain A was kept fixed, domain B was superimposed over the closest symmetry mate in the linker region. The molecular replacement could not be carried out with the extended molecule (Protein Data Bank code 3ezm) due to the relative reorientation of the two domains under different conditions. The CNS (27) maximum likelihood refinement procedure was used, combined with simulated annealing. The model was rebuilt into density using O (28) with the oligosaccharide models generated with InsightII (Accelrys) and parameterized using XPLO-2D (29).

RESULTS AND DISCUSSION

The Overall Structure of CV-N—Two different crystal forms of CV-N have been described. Trigonal crystals grow predominantly at low pH (4.6) (15), whereas tetragonal crystals grow at high pH (9.5–10.3) (19). The structures of the pseudo-monomers of CV-N seen in both crystal forms are very similar (Fig. 2A), but their relative orientation is quite different (19), and the intermolecular contacts are distinct. Only the tetragonal crystals could be used successfully for the determination of the structures of oligosaccharide complexes.

Oligosaccharides and Their Binding—Two branched high mannose oligosaccharides were used in this study. Man-9 was derived from natural sources, and its structure corresponds to an oligosaccharide that is part of the HIV-1 envelope glycoprotein gp120. The structure (Fig. 1A) is made up of a reducing end consisting of two $\beta1{\rightarrow}4$-linked N-acetyl glucosamine residues that form a chitobiose unit followed by 9 mannose residues that form a branching, triantennary structure. The second oligosaccharide (Fig. 1B) is a synthetic structure with six mannoses linked similarly to the core structure of Man-9, excluding the chitobiose unit on the reducing end (replaced by a pentyl group) and the terminal mannose units from the triantennary arms of Man-9.

CV-N has been reported to bind to Man-9 with nanomolar (13–20 nM) affinity (10, 13) and to a synthetic hexamannoside with low micromolar (2.6 μM) affinity (13). There are two distinct sugar-binding sites in a CV-N monomer, located ~35 Å away from each other: a high affinity primary site and a low affinity secondary site (10). These sites were mapped to the surface of CV-N by NMR perturbation studies even before any direct structural data on oligosaccharide binding became available. Residues 41–44, 50–56, and 74–78 define the primary site, whereas residues 1–7, 22–26, and 92–95 form the secondary site (10). In a domain-swapped dimer, there are four sugar-binding sites: two primary sites near the hinge region and two secondary sites on the opposite sides of the dimer, where it is not influenced by the conformation of the hinge region. In the domain-swapped dimer, the sugar-binding sites are formed by intertwined loops of residues belonging to both monomers. For example, the secondary site is created by residues 1–7 and 22–26 from the first monomer and residues 92'–95' and 101' from the second monomer.

The primary sugar-binding site consists of a deep pocket in the close proximity of the hinge region, which can accommodate an $\alpha1{\rightarrow}2$-linked dimannose molecule (30) (Fig. 2B). In our structures (Table I), there was no sugar bound in the primary site over a wide pH range, regardless of crystal soaking or co-crystallization with the oligosaccharide. Instead, there was a well defined, tightly bound CHES molecule from the crystallization solution, partially obstructing the pocket and forming a strong hydrogen bond anchoring the sulfate oxygen atom O13 to Arg-76 NH_2 of CV-N. Further interactions with CV-N were mainly hydrophobic, through the cyclohexyl ring of CHES. The buffer molecule did not bind the secondary sugar-binding site, even in the absence of any sugar. As crystals can be grown in other buffers in the same pH range (CAPS, CAPSO, AMP, glycine, and ethanolamine), it was unlikely that the CHES molecule was involved in any critical packing interactions. We assessed the binding affinity of CHES to CV-N using the technique of ITC (Table II). We have used ITC previously to characterize the binding interactions of various oligomannoses to CV-N (13). The binding interaction between CHES and CV-N was very weak (K_d value of ~0.1–0.5 mM) and characterized by few polar/electrostatic interactions (ΔH value of ~0.2 kcal/mol). This was unlike the oligomannose-CV-N interactions in which strong favorable binding enthalpies (ΔH value of ~−20 kcal/mol) had resulted in submicromolar CV-N binding affinities (K_d 0.02–50 μM) for the oligomannoses (13). Given these calorimetric data, it is highly unlikely that CHES could have prevented oligosaccharide binding.

Comparison of crystal (15, 19) and solution (14, 30) structures of CV-N reveals the changing geometry of the primary

sugar-binding site upon domain swapping (Fig. 3). We can only speculate on the significance, if any, of this shift in the relative orientation of the two domains. Since the primary sugar-binding site is in close proximity to the hinge region, the position of the hinge and relative orientation of the domains has a direct impact on the shape of the primary sugar-binding site. In the

monomer, the pocket is intact, accommodating a disaccharide in a stacked conformation (30), whereas in all domain-swapped structures, some of the essential protein-oligosaccharide hydrogen bonds cannot be established, leaving only the shape
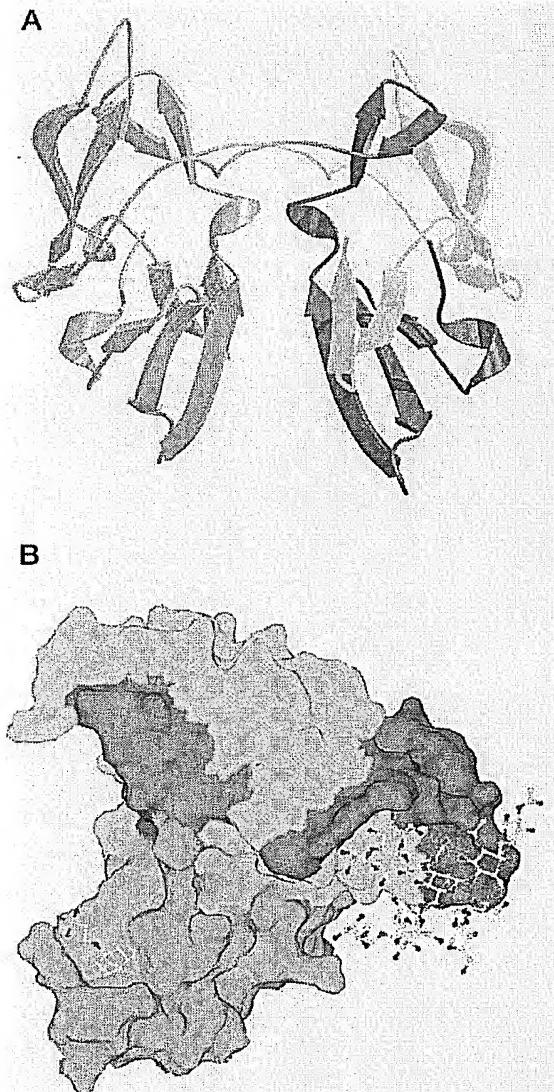
FIG. 2. **The domain-swapped CV-N dimer.** *A*, dimer with the sequence-defined domains A (*dark blue*) and B (*light blue*) from the first molecule and domains A′ (*orange*) and B′ (*yellow*) from the second molecule. AB′ and A′B are pseudo-monomers. Molecular surface (*B*) with the primary and secondary oligosaccharide-binding sites with CHES bound to the former and Man-9 to the latter. Both parts of this figure were generated with programs Bobscript (39), Raster3D (40), and SPOCK (41).
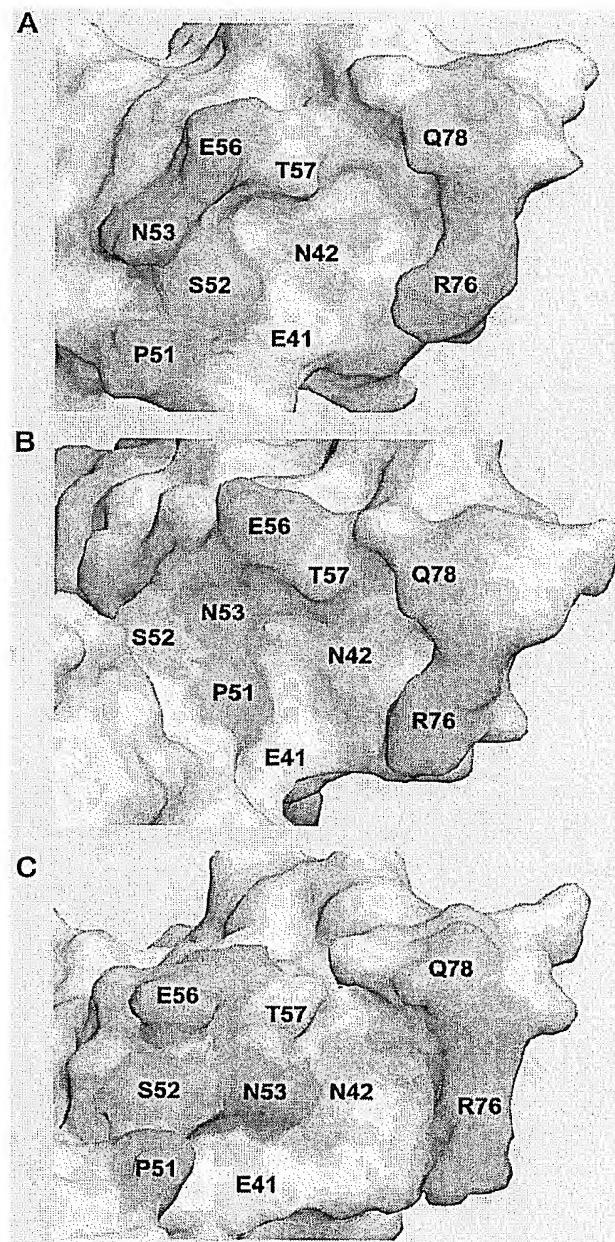
FIG. 3. **Primary oligosaccharide-binding site architectures are shown in the compact monomer (*A*), the low pH domain-swapped structure (*B*), and the high pH domain-swapped structure (*C*).**

TABLE II
Isothermal titration calorimetry results for CHES binding

	ΔH	TΔS	ΔG[a]	K_d	N (CHES:CV-N)
	kcal/mol	*kcal/mol*	*kcal/mol*	*mM*	
CHES	0.156 ± 0.009	5.036 ± 0.511	−4.881 ± 0.511	0.3 ± 0.2	1.00

[a] $\Delta G = -RT \ln K_n$; T = 303 K.

Cyanovirin-N-Sugar Complexes
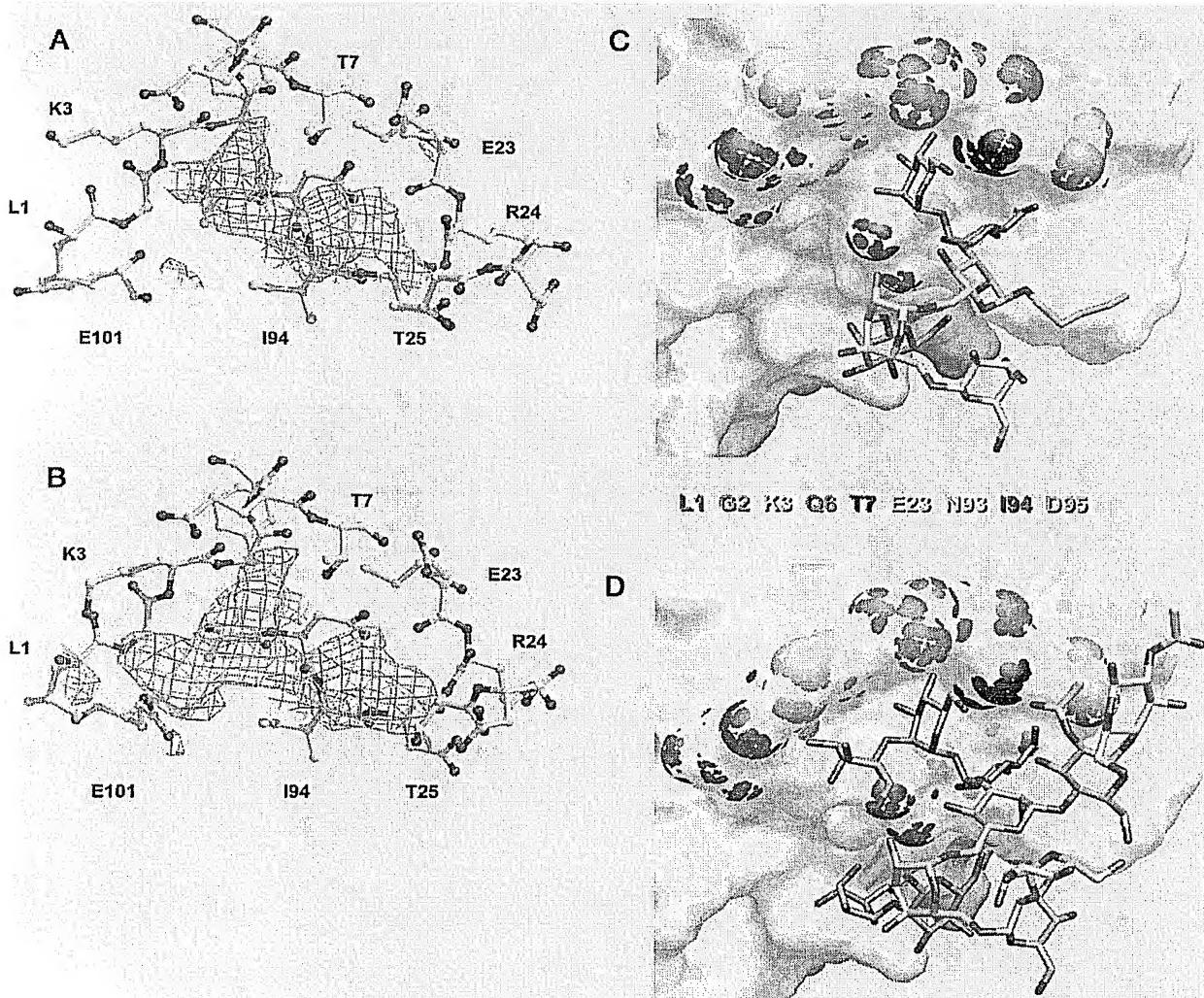
FIG. 4. **Unbiased $F_o - F_c$ difference electron density maps of the binding interfaces contoured at 2.0-σ level and the atomic models of bound oligosaccharides.** *A*, the density for the complex with hexamannoside, visible for two stacked mannose rings: C (residue 505) and 4 (residue 504). *B*, the density for the complex with Man-9, visible for three stacked mannose rings: D1 (residue 506), C (residue 505), and 4 (residue 504). The final atomic coordinates of the oligosaccharides and the protein atoms in contact with them are shown for hexamannoside (*C*) and for Man-9 (*D*). *C* and *D* were generated with the program GRASP (42).

complementarity of oligosaccharide and pocket. Specifically, in the low pH (trigonal) crystal structure (15), the sugar-binding pocket as seen in the compact monomeric structures (Fig. 3*A*) is partially changed. The OG oxygen of Ser-52 is displaced from its optimal position in which it provides a hydrogen bond to an oxygen atom from a mannose ring, disrupting one of the important hydrogen bonds and making sterically unfavorable the binding of any saccharide (Fig. 3*B*). This is the case in all trigonal crystal structures at pH 4.6–8.5, which adopt the same domain orientation and do not show oligosaccharide binding in the primary site. The high pH (tetragonal) crystal structures have a different relative orientation of the domains but still show a perturbed binding pocket. In this case, Asn-53 from the hinge region moves farther into the pocket, reducing its volume (Fig. 3*C*) and introducing unfavorable geometry for the specific binding of a mannose ring. Because of these steric considerations, sugar binding in this pocket may not be favorable and was not observed in any of the tetragonal crystal

structures at various pH values. Since we do not observe binding into the primary sugar-binding site, the question arises whether the binding of Man-9 into this pocket is also carried out by three stacked rings, as in the secondary binding site, or by two stacked rings only, as reported previously for the dimannoside (30).

The secondary sugar-binding site is located ~35 Å away from the primary site and is not affected by the geometry of the hinge, presenting the same conformation both in monomeric and domain-swapped dimeric CV-N. Comparing the two structures, the stacked position of the rings is evident from the clear electron density. Three rings are seen in the case of Man-9 (Fig. 4*A*), and two are seen in the case of hexamannoside (Fig. 4*B*). Inspection of the Man-9 structure obtained by molecular dynamics (31) reveals that branches D1 and D3 are more accessible for extended interaction than arm D2. The three terminal rings from both arm D1 and arm D3 were modeled into the electron density and refined. Due to the $\alpha1\rightarrow6$ linkage, the

TABLE III
Oligosaccharide–protein hydrogen bonds

Man-9	Ring name[a]	CV-N	Distance
			Å
504 O3	4	B Asp-95 N	3.2
504 O4	4	A Thr-25 N	2.92
504 O6	4	A Glu-23 OE2	3.2
505 O3	C	A Lys-3 N	3.54
505 O3	C	B Asn-93 N	2.67
505 O4	C	A Thr-7 N	3.52
505 O4	C	A Thr-7 OG1	2.87
506 O3	D1	A Gly-2 N	3.13
506 O3	D1	A Lys-3 NZ	3.16
506 O4	D1	A Lys-3 NZ	3.66
Hexamannoside			
504 O3	4	B Asp-95 N	2.9
504 O4	4	B Asp-95 N	3.03
504 O4	4	B Gly-96 N	3.62
504 O4	4	A Thr-25 N	3.57
504 O6	4	A Glu-23 OE2	3.16
505 O3	C	B Asn-93 N	2.79
505 O3	C	A Lys-3 N	3.4
505 O4	C	A Thr-7 N	3.63
505 O4	C	A Thr-7 OG1	3.09

[a] According to nomenclature from Fig. 1.

TABLE IV
Oligosaccharide torsion angles for α1 → 2 linkages

Linkage between rings	Oligosaccharide	ϕ	ψ
		degrees	*degrees*
C → 4	Man-9	−4.5	22.7
C → 4	hexamannoside	−3.6	10.7
D1 → C	Man-9	−39.4	45.9
2 → 1	dimannoside[a]	−48.8	36.9
2 → 1	methyl-dimannoside[b]	−55.5	14.5

[a] From CV-N/dimannoside solution structure (1iiy.pdb) (30).
[b] From methyl-dimannoside crystal structure (34).

geometry of arm D3 is not compatible with the observed electron density; only mannose rings in an α1→2-linked conformation can fit into the Man-9 electron density map. This observation is in agreement with previous results that show that the oligosaccharides interface with CV-N via a branch containing mainly α1→2-linked rings. The position of mannose rings C and 4 (residues 505 and 504) from arm D1 is very similar to the model suggested from the solution structure of a CV-N disaccharide complex (30). However, the stacked rings interact much more tightly with the protein than suggested by the NMR model, making 10 hydrogen bonds for the Man-9 and 9 hydrogen bonds for the hexamannoside binding interface (Table III). Seven of these hydrogen bonds are conserved between the binding rings 4 and C of the two oligosaccharides, but they are not comparable with the dimannoside hydrogen bonds. Mannose ring 4 of the hexamannoside with five sugar-protein hydrogen-bonds interact the most tightly. However, the higher number of hydrogen bonds for rings 4 and C of hexamannoside does not provide extra affinity as compared with Man-9. In this respect, the terminal ring D1 (residue 506) of Man-9 must be responsible for the orders of magnitude higher affinity of this ligand. The D1 ring forms three hydrogen bonds with Gly-2 and Lys-3 and could also interact with the flexible Glu-101 from domain B. These important interactions of the terminal ring D1 were not accounted for in a recent mutagenesis study of the secondary sugar-binding site using a two-ring binding interface model (32). The branching mannose ring (residue 503) can also be identified in both maps, suggesting that the α1→2-linked stacked rings are of the main D1 arm. Since the only difference between the D1 branches in Man-9 and hexamannoside is the lack of one α1→2-linked mannose ring (residue 506) at the end of the branch in the latter, the density unequivocally identifies the D1 arm as the binding interface.

The electron density was very clear for the binding interface (residues 504–506) and the branching ring 3 (residue 503) but is less well defined for the rest of the sugar molecule, which is pointing away from the protein. The well defined rings (residues 504–506) were modeled and refined first. These mannopyranose rings adopt the chair conformation with the nonreducing pyranose ring stacked over the reducing mannopyranose. NMR and computational studies show that the conformations of disaccharide fragments of an oligosaccharide fall within the allowed regions of the corresponding disaccharide φ and ψ

angle maps (33). If the pyranose rings are treated as flexible during the energy minimization studies of disaccharides, then the φ and ψ conformational space accessed is larger than when the pyranose rings are treated as rigid (33). This seems to be the case of our structures, in which φ and ψ values show more variation, possibly because of the flexibility of the mannopyranose rings. Comparison of the torsion angles for the α1→2 linkages present in the binding interface (Table IV) shows a rather good agreement of the values of φ and ψ for the terminal D1 → C ring linkage in Man-9 with the dimannoside solution structure and the methyl-dimannoside crystal structure (34). The values of φ and ψ are in good agreement for the C → 4 ring linkage in Man-9 and hexamannoside. However, φ shows a 35–50° deviation, and ψ a 5–35° deviation, when compared with the D1 → C linkage in Man-9 and the dimannoside solution structure. Since ring D1 is absent in the hexamannoside, this difference in the torsion angles could be attributed to the steric constraints of the flanking mannopyranoside ring 3 (residue 503) and those of the protein-binding site. The flexible D2 and D3 arms, which point to the solution, can adopt multiple conformations without steric strain on the binding interface. Of the two identical secondary binding sites, one had better electron density. In this site, the bound Man-9 molecule reaches over to a symmetry-related CV-N molecule, contacting residues 15, 59, and 61. These additional interactions might explain the better anchoring of the Man-9 molecule in this site and therefore its better electron density. In the other site, only the D1 arm binding rings and the branching Man-3 were modeled in and refined.

The GlcNAc ring 1 (residue 501) from Man-9 is pointing into the solvent, its most remote atom being located ~14 Å from the CV-N surface. Since this distance was measured in the most extended conformation of Man-9 bound to the secondary binding site, closer positioning of the GlcNAc ring might be possible given the flexibility of the two intermediate rings. GlcNAc 1 is anchored by an Asn residue on the surface of gp120, positioning bound gp120 much closer to CV-N than suggested previously (10). Therefore, besides oligosaccharide-mediated interactions, discrete protein-protein interactions might also play a role in the gp120-CV-N binding (12).

Several co-crystallization conditions were identified for CV-N with hexamannoside. However, the co-crystal structures did not show any sugar binding. This lack of hexamannoside binding to the secondary binding site of CV-N is similar to the results of solution phase NMR experiments in which binding of the hexamannoside to the secondary binding site was not detected (13). Also, soaks with a synthetic nonamannoside and trimannoside did not yield binding in their respective crystal structures. In both synthetic oligosaccharides, a short hydrophobic pentyl segment substituted GlcNAc 1 and 2. Nonamannoside had a tendency to disorder the crystals, raising the mosaicity significantly. The Man-9-CV-N complex structure shows that the branching Man-3 is already protruding from the binding site, interacting partially with the flexible loop region

(residues 22–27). The position of the pentyl is further restricted in nonamannoside by the bulk of the longer branches and might not interact favorably with the flexible loop region, lowering considerably the binding affinity of the sugar and at the same time having a chaotropic effect on the ordered molecules in the crystal.

CV-N undergoes conformational changes upon binding of gp120 and gp41 (6, 9, 35) with an average 11% loss of β-sheet and 2% loss of helical structure. We attempted to determine whether this reflects in the flexibility of the loop in the uncomplexed and sugar-bound forms of CV-N by comparing the B-factors of these structures. There is a well conserved trend of higher than average B-factors in the regions with residues 20–30, 50–57, and 70–80 (data not shown). Not surprisingly, these are residues directly involved in the architecture of sugar-binding sites. This suggests a flexible pocket architecture, which is able to accommodate two or three stacked rings, depending on the structure of the oligosaccharide.

Implications of the Structures for gp120 Binding—The crystallographic data presented here give visual definition to the molecular basis of the unique specificity of CV-N for Man-8 and Man-9 moieties. The additional binding affinity of CV-N for Man-8 and Man-9 as compared with Man-6 or hexamannoside is the result of the additional binding energy derived from the third mannose ring. The physiological significance of this difference is manifested in the relative paucity of Man-8 and Man-9 residues on normal mammalian glycoproteins. Man-8 and Man-9 oligosaccharides are rarely found in the human system and then usually only on glycoproteins destined for short term use and rapid degradation (*e.g.* tissue plasminogen activator). In the case of certain viral glycoproteins, however, Man-8 and Man-9 oligosaccharides are more prevalent. Such is the case of HIV gp120, where as many as 11 high mannose oligosaccharides can be present with five to six of these generally in the form of Man-8 or Man-9 (36). It has been shown that CV-N inhibits the binding of the broadly neutralizing anti-HIV-1 antibody 2G12 (6). Incubation of gp120 with excess CV-N prevents subsequent binding of 2G12 to gp120; however, CV-N binding to gp120 that was preincubated with excess 2G12 was only slightly reduced as compared with binding to gp120 alone (37). These data suggest that 2G12 binds only a subset of the Man1α→2Man termini present on gp120, whereas CV-N binds essentially all such residues (37). The ability of CV-N to target these virus-associated oligosaccharides with high affinity, whereas only binding with relatively low affinity to other, more common, mammalian oligosaccharides (*e.g.* Man-6) is the basis for the potential utility of this agent as an anti-HIV microbicide.

REFERENCES

1. Piot, P. (1998) *Science* 280, 1844–1845
2. Burton, D. R. (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 10018–10023
3. Lange, J. M., Karam, M., and Piot, P. (1993) *Lancet* 342, 1356
4. Boyd, M. R., Gustafson, K. R., McMahon, J. B., Shoemaker, R. H., O'Keefe, B. R., Mori, T., Gulakowski, R. J., Wu, L., Rivera, M. I., Laurencot, C. M., Currens, M. J., Cardellina, J. H., Buckheit, R. W., Jr., Nara, P. L., Pannell, L. K., Sowder, R. C., and Henderson, L. E. (1997) *Antimicrob. Agents Chemother.* 41, 1521–1530
5. Dey, B., Lerner, D. L., Lusso, P., Boyd, M. R., Elder, J. H., and Berger, E. A. (2000) *J. Virol.* 74, 4562–4569
6. Esser, M. T., Mori, T., Mondor, I., Sattentau, Q. J., Dey, B., Berger, E. A., Boyd, M. R., and Lifson, J. D. (1999) *J. Virol.* 73, 4360–4371
7. Geyer, H., Holschbach, C., Hunsmann, G., and Schneider, J. (1988) *J. Biol. Chem.* 263, 11760–11767
8. Leonard, C. K., Spellman, M. W., Riddle, L., Harris, R. J., Thomas, J. N., and Gregory, T. J. (1990) *J. Biol. Chem.* 265, 10373–10382
9. O'Keefe, B. R., Shenoy, S. R., Xie, D., Zhang, W., Muschik, J. M., Currens, M. J., Chaiken, I., and Boyd, M. R. (2000) *Mol. Pharmacol.* 58, 982–992
10. Bewley, C. A., and Otero-Quintero, S. (2001) *J. Am. Chem. Soc.* 123, 3892–3902
11. Bolmstedt, A. J., O'Keefe, B. R., Shenoy, S. R., McMahon, J. B., and Boyd, M. R. (2001) *Mol. Pharmacol.* 59, 949–954
12. Shenoy, S. R., O'Keefe, B. R., Bolmstedt, A. J., Cartner, L. K., and Boyd, M. R. (2001) *J. Pharmacol. Exp. Ther.* 297, 704–710
13. Shenoy, S. R., Barrientos, L. G., Ratner, D. M., O'Keefe, B. R., Seeberger, P. H., Gronenborn, A. M., and Boyd, M. R. (2002) *Chem. Biol. (Lond.)*, in press
14. Bewley, C. A., Gustafson, K. R., Boyd, M. R., Covell, D. G., Bax, A., Clore, G. M., and Gronenborn, A. M. (1998) *Nat. Struct. Biol.* 5, 571–578
15. Yang, F., Bewley, C. A., Louis, J. M., Gustafson, K. R., Boyd, M. R., Gronenborn, A. M., Clore, G. M., and Wlodawer, A. (1999) *J. Mol. Biol.* 288, 403–412
16. Bewley C. A., and Clore, G. M. (2000) *J. Am. Chem. Soc.* 122, 6009–6016
17. Barrientos, L. G., Campos-Olivas, R., Louis, J. M., Fiser, A., Sali, A., and Gronenborn, A. M. (2001) *J. Biomol. NMR* 19, 289–290
18. Clore, G. M., and Bewley, C. A. (2002) *J. Magn. Reson.* 154, 329–335
19. Barrientos, L., Louis, J. M., Botos, I., Mori, T., Han, Z., O'Keefe, B. R., Boyd, M. R., Wlodawer, A., and Gronenborn, A. (2002) *Structure* 10, 673–686
20. Schlunegger, M. P., Bennett, M. J., and Eisenberg, D. (1997) *Adv. Protein Chem.* 50, 61–122
21. Gustafson, K. R., Sowder, R. C., Henderson, L. E., Cardellina, J. H., McMahon, J. B., Rajamani, U., Pannell, L. K., and Boyd, M. R. (1997) *Biochem. Biophys. Res. Commun.* 238, 223–228
22. Botos, I., Mori, T., Cartner, L. K., Boyd, M. R., and Wlodawer, A. (2002) *Biochem. Biophys. Res. Commun.* 294, 184–190
23. Ratner, D. M., Plante, O. J., and Seeberger, P. H. (2002) *Eur. J. Org. Chem.* 2002, 826–833
24. Mori, T., Gustafson, K. R., Pannell, L. K., Shoemaker, R. H., Wu, L., McMahon, J. B., and Boyd, M. R. (1998) *Protein Expression Purif.* 12, 151–158
25. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* 276, 307–326
26. Navaza, J. (1994) *Acta Crystallogr. Sect. A* 50, 157–163
27. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 54, 905–921
28. Jones, T. A., and Kjeldgaard, M. (1997) *Methods Enzymol.* 277, 173–208
29. Kleywegt, G. J., and Jones, T. A. (1997) *Methods Enzymol.* 277, 208–230
30. Bewley, C. A. (2001) *Structure* 9, 931–940
31. Woods, R. J., Pathiaseril, A., Wormald, M. R., Edge, C. J., and Dwek, R. A. (1998) *Eur. J. Biochem.* 258, 372–386
32. Chang, L. C., and Bewley, C. A. (2002) *J. Mol. Biol.* 318, 1–8
33. Rao, V. S. R., Qasba, P. K., Balaji P. V., and Chandrasekaran, R. (1998) *Conformation of Carbohydrates*, Harwood Academic Publishers, Amsterdam, The Netherlands
34. Srikrishnan, T., Chowdhary, M. S., and Matta, K. L. (1989) *Carbohydr. Res.* 186, 167–175
35. Mariner, J. M., McMahon, J. B., O'Keefe, B. R., Nagashima, K., and Boyd, M. R. (1998) *Biochem. Biophys. Res. Commun.* 248, 841–845
36. Yeh, J. C., Seals, J. R., Murphy, C. I., van Halbeek, H., and Cummings, R. D. (1993) *Biochemistry* 32, 11087–11099
37. Scanlan, C. N., Pantophlet, R., Wormald, M. R., Ollmann, S. E., Stanfield, R., Wilson, I. A., Katinger, H., Dwek, R. A., Rudd, P. M., and Burton, D. R. (2002) *J. Virol.* 76, 7306–7321
38. Brünger, A. T. (1992) *Nature* 355, 472–474
39. Esnouf, R. M. (1997) *J. Mol. Graph. Model.* 15, 132–134
40. Meritt, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 50, 869–873
41. Christopher, J. A. (1998) *SPOCK: The Structural Properties Observation and Calculation Kit*, Version 1.06170, The Center for Macromolecular Design, Texas A&M University, College Station, TX
42. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins Struct. Funct. Genet.* 11, 281–296

Solution structure of cyanovirin-N, a potent HIV-inactivating protein

Carole A. Bewley[1], Kirk R. Gustafson[2], Michael R. Boyd[2], David G. Covell[3], Ad Bax[1], G. Marius Clore[1] and Angela M. Gronenborn[1]

The solution structure of cyanovirin-N, a potent 11,000 M_r HIV-inactivating protein that binds with high affinity and specificity to the HIV surface envelope protein gp120, has been solved by nuclear magnetic resonance spectroscopy, including extensive use of dipolar couplings which provide *a priori* long range structural information. Cyanovirin-N is an elongated, largely β-sheet protein that displays internal two-fold pseudosymmetry. The two sequence repeats (residues 1–50 and 51–101) share 32% sequence identity and superimpose with a backbone atomic root-mean-square difference of 1.3 Å. The two repeats, however, do not form separate domains since the overall fold is dependent on numerous contacts between them. Rather, two symmetrically related domains are formed by strand exchange between the two repeats. Analysis of surface hydrophobic clusters suggests the location of potential binding sites for protein–protein interactions.

The initial events that lead to HIV infection include binding of the virus to the host cell, activation of the virus, and ultimately virus-cell or cell-cell fusion[1]. During the first step of HIV infection, the viral surface envelope glycoprotein gp120 interacts with the CD4 receptor of the host cell, upon which gp120 undergoes a conformational change[2] sufficient to accommodate a subsequent interaction between gp120 and a member of the α and β chemokine receptor families, now commonly referred to as coreceptors[2,3]. Concurrently, gp41 dissociates from gp120, associates with the target membrane and mediates fusion. In this paper we describe the three-dimensional solution structure of a newly discovered cyanobacterial protein, named cyanovirin-N, which is a highly potent inhibitor of diverse laboratory adapted strains and clinical isolates of HIV-1, as well as HIV-2 and SIV[4]. The antiviral activity of cyanovirin-N is mediated, at least in part, through high affinity binding to gp120[4,5]. Cyanovirin-N is currently under joint NCI/NIAID investigation as a broad-spectrum virucidal and therapeutic agent against HIV.

Cyanovirin-N was originally isolated from an aqueous extract of a cultured cyanobacterium, *Nostoc ellipsosporum*[4], and was identified in a screening effort aimed at the discovery of new sources of HIV inhibitors[6]. The primary sequence and disulfide bonding pattern were determined by conventional biochemical techniques[4,7], and a synthetic gene was constructed for over-expression of the protein[4,8]. Analysis of the primary sequence of cyanovirin-N revealed the presence of two internal repeats of 50 and 51 amino acids that show strong sequence similarity to one another, and equivalent positions of the disulfide bonds (Fig. 1)[7]. Cyanovirin-N is extremely resistant to physico-chemical degradation and can withstand treatment with denaturants, detergents, organic solvents such as acetonitrile or methanol, multiple freeze-thaw cycles, and heat (up to 100 °C) with no subsequent loss of antiviral activity[4]. The primary sequence of cyanovirin-N shares no similarity with other proteins thus far deposited in public protein data bases.

Structure determination

The solution structure of cyanovirin-N was determined using double and triple resonance multidimensional heteronuclear NMR spectroscopy, making use of uniformly ^{15}N- and ^{15}N/^{13}C-labeled protein[9–11]. The final ensemble of structures was calculated by simulated annealing[12] on the basis of 2,509 experimental NMR restraints, including 334 residual dipolar couplings ($^1D_{NH}$, $^1D_{CH}$, $^1D_{C\alpha C'}$, $^1D_{NC'}$ and $^2D_{HNC'}$). The latter are a function of the orientation of interatomic vectors relative to the molecular alignment tensor, and hence provide qualitatively different structural information from that afforded by other NMR observables, such as NOEs, coupling constants and chemical shifts, which are reliant on close spatial proximity of atoms[13]. A summary of the structural statistics is provided in Table 1, and a superposition of the final ensemble of 40 simulated annealing structures is shown in Fig. 2a. It should be noted that the inclusion of the residual dipolar couplings in the structure refinement increases the coordinate precision from 0.3 Å to 0.15 Å for the backbone atoms and from 0.54 Å to 0.45 Å for all heavy atoms. The atomic r.m.s. shift in the mean coordinates resulting from the inclusion of the dipolar couplings is 0.66 Å for the backbone atoms and 0.81 Å for all heavy atoms.

Description of the structure

Cyanovirin-N has the shape of an elongated prolate ellipsoid, ~55 Å in length with a maximum width of ~25 Å. The secondary structure elements comprise 10 β-strands and four short 3_{10}-helical turns. The overall topology can be described as follows (Figs 1a, 2b). In the first sequence repeat (residues 1–50) a four-residue helical turn (residues 3–6) precedes a three-stranded antiparallel β-sheet (β1, β2 and β3 from residues 7–14, 16–24 and 28–36 respectively). β-strand 1 is characterized by a wide β-bulge[14] at Asn 10 and Ser 11. A three-residue helical turn (residues 37–39) connects β-strand 3 to a β-hairpin formed by β-strands 4 (residues 39–43) and 5

articles

a
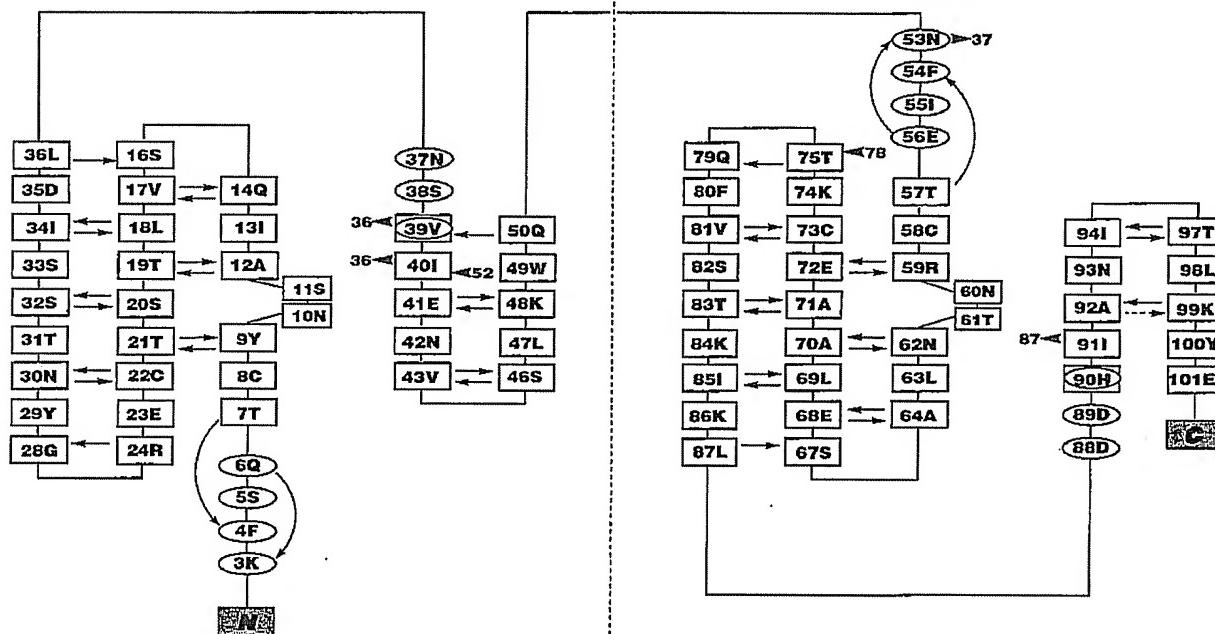
b
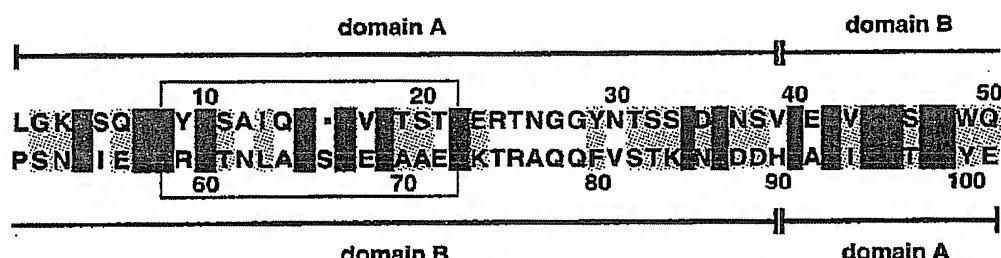
Fig. 1 Secondary structure elements and structure alignment of cyanovirin-N. *a*, Secondary structure and global topology of cyanovirin. Boxed residues denote β-strands, circled residues 3₁₀-helical turns, solid arrows indicate backbone hydrogen bonds from amide protons to carbonyl oxygens, the dotted arrow indicates hydrogen bonds bridged by a water molecule, and offset boxes denote the positions of β-bulges. The dotted line demarcates the first and second sequence repeats. *b*, Structure alignment of the first and second sequence repeats. Yellow boxes denote conserved residues, red boxes identical residues; black and red lines demarcate domains A and B respectively.

(residues 46–50). The β-hairpin is directed away from the three-stranded antiparallel β-sheet at an angle of ~140° (Figs 2*b*, 3*a*). This topology is exactly repeated in the second sequence repeat (residues 51–101) wherein the second triple-stranded antiparallel β-sheet is formed by β-strands 6 (residues 57–64 with a wide β-bulge at Asn 60 and Thr 61), 7 (residues 67–75) and 8 (residues 79–87), and the second β-hairpin by β-strands 9 (residues 91–94) and 10 (residues 97–100). The first and second sequence repeats are oriented opposite to one another with respect to the pseudosymmetric two-fold axis, which is directed into the plane of the paper in the view shown in Fig. 2*b*, such that β-strands 1 and 6 lie adjacent and are oriented antiparallel to one another. This arrangement places each of the β-hairpins on top of the triple stranded β-sheet of the other half of the molecule with an angle of ~40° between the β-hairpin and the central strand of the underlying β-sheet.

Although cyanovirin-N can be clearly divided into two sequential sequence repeats, the individual repeats do not form separate domains (Fig. 3*a*,*b*). Rather, the overall fold of

the molecule depends on numerous contacts between the two repeats. Indeed, the interaction between the two repeats of cyanovirin-N buries 3,085 Å² of accessible surface area (1,551 Å² for the first repeat and 1,534 Å² for the second). In addition, there are several electrostatic interactions between the first and second repeats which include hydrogen bonds between Gly 15(NH) and Thr 61(Oγ), Asn 37(NδH₂) and Asn 53(Oδ), Asp 44(Oδ) and both Arg 76(NH) and Arg 76(Nε), and Leu 47(NH) and Thr 83(Oγ).

The overall structure of cyanovirin-N, however, can be divided into two symmetrically related domains, A and B, formed by strand exchange between the two sequence repeats (Fig. 3*c*). Domain A contains the N- and C-termini and comprises residues 1–39 and 90–101; domain B extends from residues 39 to 90. Thus, domains A and B correspond to the 'top' and 'bottom' halves of the molecule in the views shown in Fig. 3*a*,*c*. Each domain contains the triple-stranded antiparallel β-sheet of one repeat and the β-hairpin of the other repeat, and the two domains are joined together by helical turns 2 (residues 37–39) and 4 (residues 88–90) (Fig. 3*c*).
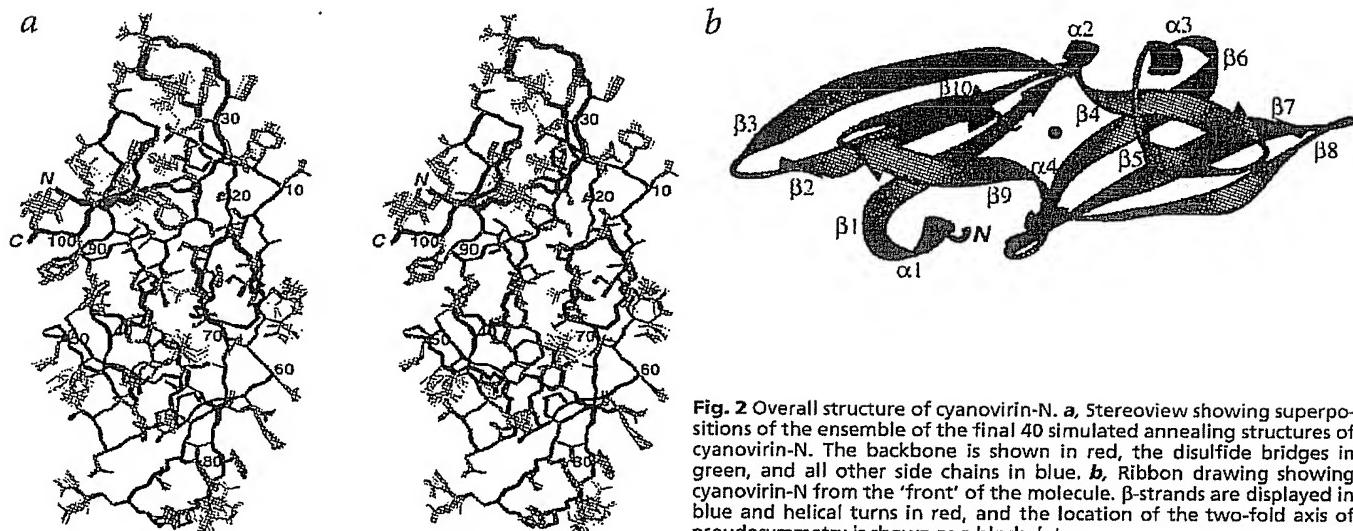
Fig. 2 Overall structure of cyanovirin-N. **a,** Stereoview showing superpositions of the ensemble of the final 40 simulated annealing structures of cyanovirin-N. The backbone is shown in red, the disulfide bridges in green, and all other side chains in blue. **b,** Ribbon drawing showing cyanovirin-N from the 'front' of the molecule. β-strands are displayed in blue and helical turns in red, and the location of the two-fold axis of pseudosymmetry is shown as a black dot.

The positioning of the β-hairpins with respect to the underlying triple-stranded β-sheets is determined by numerous hydrophobic interactions. In domain A, Phe 4, Leu 18, Ile 34 and Leu 36 located in the β-sheet of the first repeat interact with Leu 87, Ile 91 and Leu 98 which are part of the β-hairpin of the second repeat (Fig. 4a). Note that leucines 36 and 87 are located close to the two-fold axis and assume inverted positions in the packing of domain B (Fig. 4b). The disulfide bridge between Cys 8 and Cys 22 has a right-handed hook conformation, and the Sγ atoms are in van der Waals contact with side chains of Phe 4, Asn 93 and Leu 98. A hydrogen bond between the NδH₂ group of Asn 93 and the Oγ atom of Thr 7 bridges β-strands 9 and 1. Finally, a tightly bound water molecule, identified by the observation of ROEs from water to the amide protons of Ala 92, Lys 99 and Glu 101, serves to bridge hydrogen bonds between the backbone carbonyls of His 90 and Lys 99 and the backbone amide of Ala 92.

Fig. 4b shows domain B viewed in the same orientation as domain A displayed in Fig. 4a. A nearly identical set of interactions is apparent: namely, the equivalently positioned side chains of Phe 54, Leu 69, Ile 85 and Leu 87 of the second triple-stranded β-sheet pack against Leu 36, Ile 40 and Leu 47 of the first β-hairpin. Note that the side chain conformation of Ile 85 with χ_1/χ_2 angles in the t/t rotamers differs from the equivalent Ile 34 which has χ_1/χ_2 angles in the g⁻/t rotamers, presumably due to the presence of the tightly bound water wedged between β-strands 9 and 10. The disulfide bridge between Cys 58 and Cys 73 occupies the same position as that between Cys 8 and Cys 22, with the Sγ atoms in close contact with Phe 54, Asn 42 and Leu 47; and a hydrogen bond occurs between the side chains of Asn 42 (in β-strand 4) and Thr 57 (in β-strand 6). In addition, the side chains of Thr 61 and Ala 71, which are in van der Waals contact with one another, pack against the aromatic ring of Phe 54 providing an auxiliary element to the core not seen in domain A where two serines (at positions 11 and 20) are substituted for Thr 61 and Ala 71.

At the interface of the two domains (Fig. 4c), a cluster of hydrophobic residues, comprising Val 39, His 90, Trp 49 and Tyr 100, brings together helical turns 2 and 4 and the tips of β-strands 5 and 10 respectively (Fig. 4c). Finally, the interactions between Trp 49 and Asp 89, which include both hydrophobic contacts and a hydrogen bond between the NεH atom of Trp 49 and the Oδ atom of Asp 89, contribute to the closing of this gap.

Given the strong internal sequence similarity of the two repeats of cyanovirin-N, the nearly identical structures observed for the two repeats is not surprising. A structural alignment between the first and second repeats show 16 residues to be identical and 19 residues to be conservatively replaced giving rise to an overall 70% similarity (Fig. 1b), and the two repeats can be superimposed with a Cα atomic r.m.s. difference of 1.3 Å for all 50 residues (Fig. 3b). Likewise, domains A and B can be superimposed with a Cα atomic r.m.s. difference of 1.3 Å (Fig. 3d). A noteworthy feature that results from the internal two-fold pseudosymmetry is the adjacent placement of the N- and C-termini, wherein the side chains of Leu 1 and Glu 101 are in van der Waals contact. Thus Leu 1 and Glu 101 form a noncovalent bridge over β-strand 9 that mimics the loop connecting β-strand 5 and helical turn 3, which similarly crosses over β-strand 4 (Fig. 2b). This close arrangement of the N- and C-termini may help to explain results from mutagenesis studies that showed that removal of three consecutive residues from either the N- or C-terminus reduces the antiviral activity by over two orders of magnitude[5].

Similarities to other proteins

The function of cyanovirin-N in the cyanobacterium from which it was isolated remains unknown. An automated search of protein and gene sequence data banks[15] failed to return any sequences similar to cyanovirin-N. Similarly, a search of the Brookhaven protein structure database using the program DALI[16] did not reveal any proteins with significant overall structural similarity (Z score > 3) to cyanovirin-N. The two individual domains, however, bear a distant resemblance to the SH3 fold[17] (Fig. 5). The closest match is with the hyperthermophile DNA binding protein Sac7d[18] for which 41 residues can be superimposed onto each domain of cyanovirin-N with a Cα atomic r.m.s. difference of ~3 Å. The match comprises the triple stranded antiparallel β-sheet and the two helical turns within each domain. The β-hairpin, however, in the domains of cyanovirin-N, is flipped by ~180° about its long axis, relative to that in Sac7d and other SH3

articles
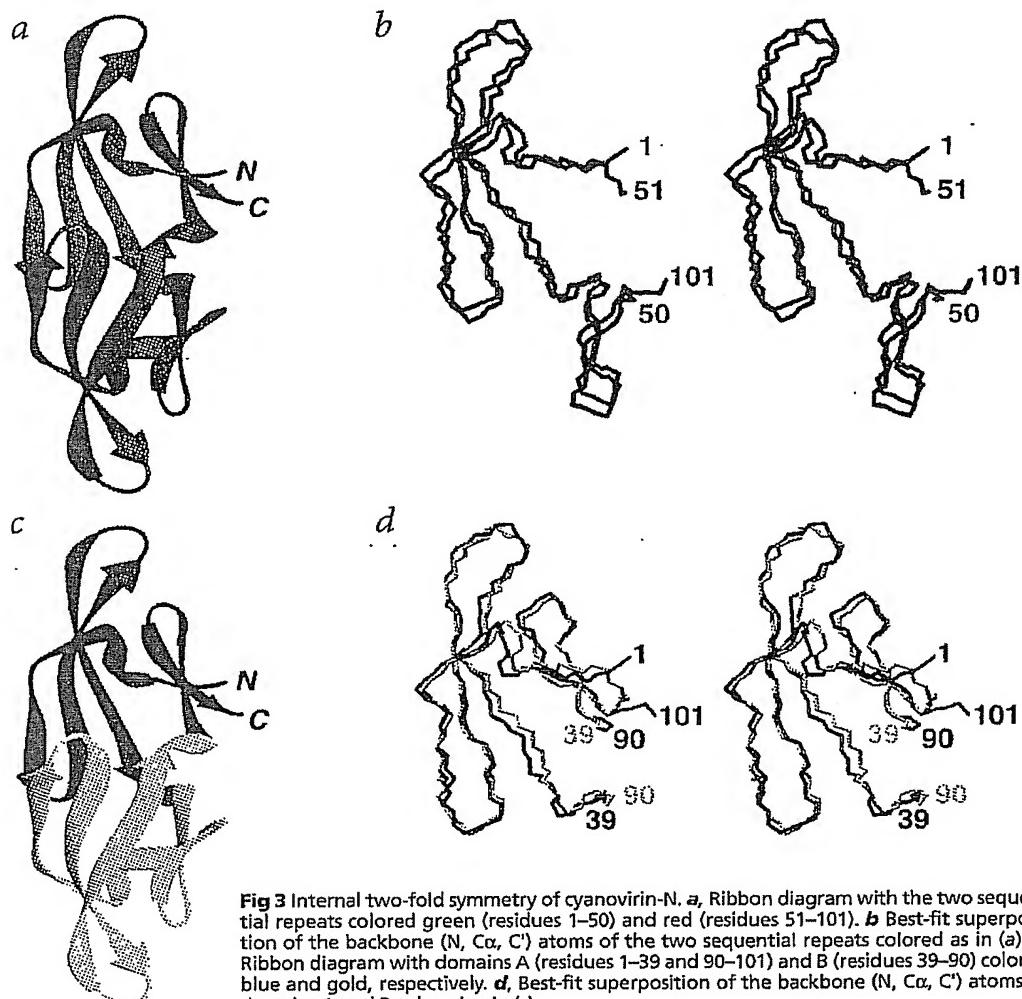
Fig 3 Internal two-fold symmetry of cyanovirin-N. *a*, Ribbon diagram with the two sequential repeats colored green (residues 1–50) and red (residues 51–101). *b* Best-fit superposition of the backbone (N, Cα, C') atoms of the two sequential repeats colored as in (*a*). *c*, Ribbon diagram with domains A (residues 1–39 and 90–101) and B (residues 39–90) colored blue and gold, respectively. *d*, Best-fit superposition of the backbone (N, Cα, C') atoms of domains A and B colored as in (*c*).

the target cell and any of its receptors (such as CD4, CXCR4, CCR5). Through a variety of experimental approaches, cyanovirin-N was shown to bind avidly to gp120, including recombinant non-glycosylated gp120[4]. Further, pretreatment of cyanovirin-N with exogenous, virus-free gp120 resulted in a concentration-dependent decrease in antiviral activity[4]. The recombinant cyanovirin-N used in the NMR structural studies had gp120 binding and anti-HIV properties that were indistinguishable from those of cyanovirin-N isolated from its natural source[1].

Since surface hydrophobicity plays a key role in protein–protein interactions[19–22], we have mapped the most hydrophobic surface clusters on cyanovirin-N using the method of Covell and coworkers[20–22] to predict which regions of cyanovirin-N may be interacting with the viral envelope. Fig. 6 shows two views of a surface representation of cyanovirin-N onto which have been mapped the electrostatic potential (left-hand panels) and the two highest ranking hydrophobic clusters (center panels). The top ranking surface hydrophobic cluster (Fig. 6*a*) is located in domain A, comprises Leu 1, Gly 2, Lys 3, Gln 6, Thr 7, Thr 25, Asn 26, Asn 93, Ile 94 and Asp 95, and is centered around the N-terminus and the turns connecting β-strands 2 and 3 and β-strands 9 and 10. This cluster of amino acids forms an extensive and curved ridge that surrounds the cleft between the first triple-stranded β-sheet and the second β-hairpin (Fig. 6*a*, right). Comparison of the hydrophobic cluster with the electrostatic surface shows that this ridge, as well as the cleft, is predominantly neutral, with the exception of the positive charge from the NζH₃ group of Lys 3 and the negative charge from the carboxylate of Asp 95 located at the top right and bottom left corners respectively, of the hydrophobic cluster. The observation that a large portion of this highest ranking hydrophobic cluster consists of the first three N-terminal residues provides a rational for the finding that a cyanovirin-N mutant lacking the first three N-terminal residues is virtually inactive[5]. It should be noted that the symmetrically equivalent surface region in domain B of cyanovirin-N, which would consist of helical turn 3 and the turns between β-strands 4 and 5 and β-strands 7 and 8 (surface not shown) carries a formal charge of -2 wherein the neutral residues Gln 6, Thr 26 and Ala 92 (not labeled) are substituted by Glu 56, Arg 76 and Glu 42

domains. It should be noted that in SH3 domains of signaling proteins the equivalent structural element to this hairpin is extended into a long, so-called RT loop which forms part of the binding site for polyproline containing peptides[17].

Potential binding surfaces of cyanovirin-N

Several observations in regard to the antiviral activity of cyanovirin-N suggested early on that its ability to inactivate diverse strains of HIV may be a result of interactions between cyanovirin-N and the HIV envelope glycoprotein gp120[4,5]. First, pretreatment of virus with cyanovirin-N prevented virus-cell fusion and infection of cells, while pretreatment of cells with cyanovirin-N offered no protection. Second, delayed addition experiments revealed that cyanovirin-N had to be added to cells before or shortly after addition of virus to afford maximum antiviral activity. And third, cyanovirin-N inactivated diverse laboratory strains and clinical isolates of HIV-1, HIV-2 and SIV, all of which share similar surface envelope glycoprotein functions. In the case of HIV-1, the clinical isolates included M-tropic, T-tropic and dual-tropic strains, all of which were inhibited at comparable low nanomolar concentrations[4]. Collectively these results suggested that cyanovirin-N inhibits fusion and viral transmission by direct interactions with the virus as opposed to

respectively, and Lys3 is replaced by Asn53, resulting in a large decrease in surface hydrophobicity and a very different charge distribution.

The second highest ranking surface hydrophobic cluster is located in domain B, comprises Ala 64, Gly 65, Ser 66, Glu 68, Ala 70, Lys 84 and Asn 86, spans regions of β-strands 6, 7 and 8 and includes the turn between β-strands 6 and 7 (Fig. 6b). Ala 70, Ala 64, Gly 65 and Ser 66 form a neutral vertical ridge while Lys 84 and Glu 68 present positive and negative charges respectively, at the left edge of the region (Fig. 6b).

It is known that cyanovirin-N does not bind to gp120 at either of the V3 loop or the CD4 binding site[4] regions of gp120, which have been well characterized. Thus, future structural studies of cyanovirin-N complexed to the relevant, folded domain of gp120 should shed light on the molecular mechanisms of the conformational changes necessary for initiation of the fusion event, and may also provide more structural clues concerning the interactions between gp120 and gp41.

Methods

Expression, purification and sample preparation. Cloning and expression of a synthetic gene for cyanovirin-N has been described elsewhere[8]. Briefly, the appropriate DNA coding sequence was subcloned into the E. coli vector pFLAG, followed by transformation of E. coli strain BL21. Uniform (>95%) ^{15}N and ^{13}C labeling was obtained by growing the cells in modified minimal medium containing $^{15}NH_4Cl$ and/or $^{13}C_6$-glucose as the sole nitrogen and carbon sources respectively. Cells were grown at 37 °C, and protein expression was induced for three hours with 1 mM isopropyl-D-thiogalactoside. The cells were harvested, resuspended in 10 mM Tris buffer, pH 7.4, 5 mM EDTA, and 5 mM benzamidine, lysed by passage through a French press, and cleared by centrifugation. The supernatant was applied directly to a preparative Bakerbond C4 wide pore column and eluted in a stepwise manner with water, 2:1 v/v water:methanol, 1:2 v/v water-methanol, and methanol. Fractions containing cyanovirin-N were combined and further purified by reversed-phase HPLC using a C18 column equilibrated with 0.05% trifluoroacetic acid and eluted with a linear gradient of 20% to 40% acetonitrile. Based on amino acid analysis and UV spectroscopy, the extinction coefficient of cyanovirin-N is 9,400 mol^{-1} cm^{-1} at 280nm. Samples for NMR contained ~1.4 mM protein at pH 6.1.

NMR spectroscopy. NMR experiments were carried out at 27 °C on Bruker DMX500, DMX600 and DMX750 spectrometers equipped with x,y,z-shielded gradient triple resonance probes. Spectra were processed with the NMRPipe package[23], and analyzed using the programs PIPP and STAPP[24]. ^1H, ^{15}N and ^{13}C resonance assignments were made from the following 3D through-bond heteronuclear correlation experiments: CBCA(CO)NH, CBCANH, HNCO, HNHA, DIPSI-H(CCO)NH, DIPSI-C(CCO)NH, HCCH-COSY and HCCH-TOCSY[9–11]. $^3J_{HN\alpha}$, $^3J_{C'C\gamma}$ (aromatic, methyl and methylene), $^3J_{NC\gamma}$ (aromatic, methyl and methylene), and $^3J_{C'C'}$ couplings were measured by quantitative J correlation spectroscopy[25]. φ and ψ backbone torsion angle restraints were derived from $^3J_{HN\alpha}$ and $^3J_{C'C_0}$ coupling constants, the three-bond amide deuterium isotope effect on the $^{13}C\alpha$ shifts (measured from a 3D HCA(CO)N recorded on a protein sample dissolved in 50% H_2O/50% D_2O)[26], and the backbone (^{15}N, NH, $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$, Hα) secondary chemical shifts using the program TALOS (G. Cornilescu, F. Delaglio and A.B., in preparation). The latter comprises a database of residue triplets correlating φ/ψ angles (derived from high resolution structures) and their corresponding backbone secondary chemical shifts, and is based on the premise that when a string of three amino acid shows high similarity in secondary shifts and residue type with a string of amino acids in the database, the central residues of the two strings are likely to have similar backbone torsion angles. Side chain torsion angles were derived from NOE/ROE data and three-bond heteronuclear coupling constants. Interproton distance restraints were derived from 3D ^{15}N (120 ms mixing time) and ^{13}C (45 and 120 ms mixing time) separated NOE experiments, and 4D $^{13}C/^{15}N$ (140 ms mixing time) and $^{13}C/^{13}C$ (140 ms
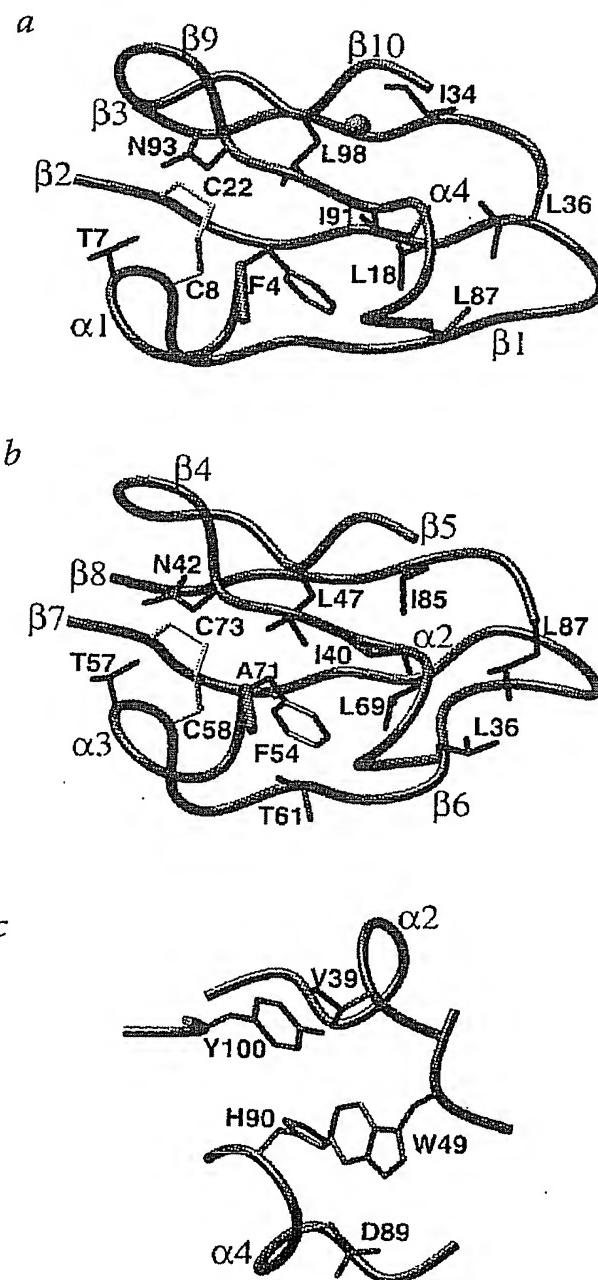
Fig. 4 Side chain contacts forming the core of cyanovirin-N. Views showing the core of **a,** domain A and **b,** domain B. Domain B is shown in the same orientation as domain A. **c,** The interface of the two domains. A Cα backbone worm is shown in blue, hydrophobic residues in red, aromatic residues in green, disulfide bonds in yellow, and all other residues in magenta. The tightly bound water in domain A which serves to bridge backbone hydrogen bonds between β-strands 9 and 10 is shown as a cyan colored sphere.

mixing time) separated NOE experiments[9–11]. Location of bound water was determined by means of a 2D H_2O-ROE-^1H-^{15}N-HSQC (mixing time 45 ms) spectrum[27]. $^1D_{NH}$, $^1D_{C\alpha H}$, $^1D_{C\alpha C'}$, $^1D_{NC'}$ and $^2D_{HNC'}$ residual dipolar couplings were obtained by taking the difference in the corresponding J splittings measured on oriented (in 4% 3:1 DMPC:DHPC at 38 °C) and isotropic (in water) cyanovirin-N[28]. $^1J_{NH}$, $^1J_{C\alpha H}$, $^1J_{C\alpha C'}$ couplings were obtained from a 2D IPAP {^{15}N,^1H}-HSQC experiment to generate two spectra containing either the upfield
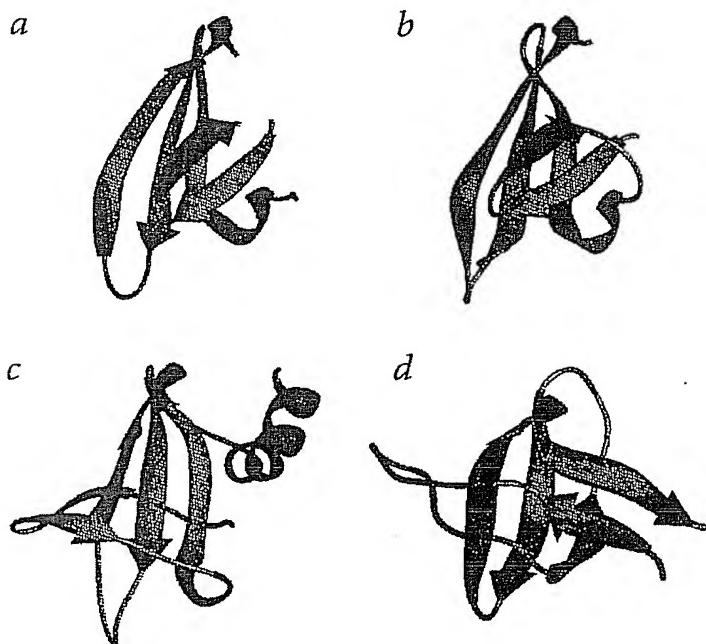
articles

The measured $^1D_{NH}$ values ranged from -31 to +20 Hz, and the normalization factors (given by $\gamma_N\gamma_H<r_{NH}^{-3}>/\gamma_A\gamma_B<r_{AB}^{-3}>$ where γ and r represent gyromagnetic ratios and distances respectively) employed for $^1D_{C\alpha H}$, $^1D_{C\alpha C'}$, $^1D_{NC'}$ and $^2D_{HNC'}$ relative to $^1D_{NH}$ were 0.48, 5.36, 9.04 and 3.04 respectively. The magnitude of the axial and rhombic components of the alignment tensor $\mathbf{D}^{NH}$ were obtained by examining the distribution of normalized dipolar couplings[31] which yielded values of D_a^{NH} = -17.0 Hz and R = 0.17, where D_a^{NH} is the axial component of the tensor and R is the rhombicity defined as the ratio of the rhombic to axial components of the tensor. This value of D_a^{NH} corresponds to a value of 1.48×10^{-3} for A_a which is the unitless axial component of the molecular alignment tensor **A**. Heteronuclear $^{15}N-\{^1H\}$ NOEs were measured as described[32] and identified only a single residue with a $^{15}N-\{^1H\}$ NOE less than 0.6, namely Ser 52 which had an NOE value of ~0.4.

Structure calculations. Approximate interproton distance restraints, derived from multidimensional NOE spectra, were grouped into four distance ranges, 1.8–2.7 Å (1.8–2.9 Å for NOEs involving NH protons), 1.8–3.3 Å (1.8–3.5 Å for NOEs involving NH protons), 1.8–5.0 and 1.8–6.0 Å, corresponding to strong, medium, weak and very weak NOEs respectively[9]. 0.5 Å was added to the upper bound for distances involving methyl groups to account for the higher apparent intensity of the methyl resonances. Distances involving non-stereospecifically assigned methylene protons, methyl groups, and Hδ and Hε protons of Tyr and Phe, were represented as a $(\Sigma r^{-6})^{-1/6}$ sum[33]. The structures were calculated by simulated annealing[12,34] using the program CNS[35], adapted to incorporate pseudopotentials for three-bond coupling constants[36], secondary $^{13}C\alpha/^{13}C\beta$ chemical shifts[37], proton chemical shifts[38,39] and residual dipolar coupling[40] restraints, and a conformational database potential for the non-bonded contacts derived from very high resolution ⟨1.7 Å or better⟩ X-ray structures[41,42]. The target function that is minimized during simulated annealing and restrained regularization comprises quadratic harmonic potential terms for covalent geometry, $^3J_{HNH\alpha}$ coupling constant restraints, secondary $^{13}C\alpha$ and $^{13}C\beta$ chemical shift restraints, ^1H chemical shift restraints, and dipolar coupling restraints; square-well quadratic potentials for the experimental distance and torsion angle restraints; a quartic van der Waals repulsion term and a conformational database potential term for the non-bonded contacts. The latter biases sampling during simulated annealing refinement to conformations that are likely to be energetically possible by effectively limiting the choices of dihedral angles to those that are known to be physically realizable[41,42]. There were no hydrogen-bonding, electrostatic, or 6–12 Lennard-Jones empirical potential energy terms in the target function.

Structure figures were generated using the programs MOLMOL[43], GRASP[44] and RIBBONS[45]. The secondary struc-

or downfield ^{15}N doublet component[29], a 3D $^1H\alpha$-coupled(F_1) HCA(CO)N experiment, and a 2D constant time $^{13}C'$-coupled/$^1H\alpha$-decoupled(F_1) ^1H-^{13}C HSQC experiment respectively. $^1J_{NC'}$ and $^2J_{HNC'}$ couplings were obtained from a 2D $^{13}C'$-coupled/$^{13}C\alpha$-decoupled(F_1) ^1H-^{15}N HSQC experiment[30]. The precision of the measured $^1D_{NH}$, $^1D_{C\alpha H}$, $^1D_{C\alpha C'}$, $^1D_{NC'}$ and $^2D_{HNC'}$ dipolar couplings was ~0.5–1.0 Hz, ~1–1.5 Hz, ~1.0–1.5 Hz, ~0.5–1.0 Hz and ~1.0–1.5 Hz respectively.

articles

Table 1 Structural statistics[1]

	<SA>	$\overline{(SA)}r$
R.m.s. deviations from experimental distance restraints (Å)[2]		
All (1,241)	0.013 ± 0.001	0.010
interresidue sequential (|i - j| = 1) (418)	0.008 ± 0.003	0.005
interresidue medium range (1 < |i - j| ≤ 5) (171)	0.005 ± 0.002	0.007
interresidue long range (1 < |i - j| ≤ 5) (540)	0.016 ± 0.002	0.014
intraresidue (20)	0.028 ± 0.011	0.016
bound water (8)	0.001 ± 0.004	0.000
H-bonds (84)	0.017 ± 0.006	0.009
R.m.s. deviations from exptl dihedral restraints (°) (334)[2]	0.266 ± 0.052	0.175
R.m.s. deviations from $^3J_{HN\alpha}$ coupling constants (Hz) (81)[2]	0.60 ± 0.01	0.61
R.m.s. deviations from secondary ^{13}C shifts (p.p.m.)		
$^{13}C\alpha$ (82)	0.85 ± 0.01	0.84
$^{13}C\beta$ (75)	1.17 ± 0.01	1.16
R.m.s. deviations from ^1H shifts (p.p.m.) (362)	0.25 ± 0.002	0.25
R.m.s. deviations from residual dipolar couplings (Hz)		
$^1D_{NH}$ (Hz) (84)	0.50 ± 0.02	0.50
$^1D_{CH}$ (Hz) (77)	1.12 ± 0.03	1.13
$^1D_{C\alpha C'}$ (Hz) (44)	1.26 ± 0.01	1.25
$^1D_{NC'}$ (Hz) (66)	0.55 ± 0.01	0.56
$^2D_{HNC'}$ (Hz) (63)	1.25 ± 0.01	1.26
Deviations from idealized covalent geometry		
bonds (Å) (1519)	0.004 ± 0.0007	0.005
angles (°) (2,724)	0.699 ± 0.007	0.757
impropers (°) (775)	0.754 ± 0.019	0.784
Measures of structure quality		
E_{LJ} (kcal mol^{-1})[3]	-434 ± 5	-434
PROCHECK[4]		
Residues in most favorable region of Ramachandran plot	85.4 ± 0.7	87.0
No. of bad contacts per 100 residues	5.7 ± 1.1	5.0
Coordinate precision (Å)[5]		
backbone (N, Cα, C', O)	0.15 ± 0.02	
all non-hydrogen atoms	0.45 ± 0.03	

[1]The notation of the NMR structures is as follows: <SA> are the final 30 simulated annealing structures; $\overline{SA}$ is the mean structure obtained by averaging the coordinates of the individual SA structures best-fitted to each other (residues 1–101); $\overline{(SA)}r$ is the restrained regularized mean structure obtained by restrained regularization of the mean structure $\overline{SA}$. The number of terms for the various restraints is given in parentheses. The final values for the force constants employed for the various terms in the target function are as follows: 1,000 kcal mol^{-1} Å^{-2} for bond lengths, 500 kcal mol^{-1} rad^{-2} for angles and improper torsions (which serve to maintain planarity and chirality), 4 kcal mol^{-1} Å^{-4} for the quartic van der Waals repulsion term (with the van der Waals radii set to 0.8 times their value used in the CHARMM PARAM19/20 parameters), 30 kcal mol^{-1} Å^{-2} for the experimental distance restraints (interproton distances and hydrogen bonds), 200 kcal mol^{-1} rad^{-2} for the torsion angle restraints, 1 kcal mol^{-1} Hz^{-2} for the $^3J_{HN\alpha}$ coupling constant restraints, 0.5 kcal mol^{-1} p.p.m.$^{-2}$ for the secondary ^{13}C chemical shift restraints, 7.5 kcal mol^{-1} p.p.m.$^{-2}$ for the ^1H chemical shift restraints, 1.0 kcal mol^{-1} Hz^{-2} for the $^1D_{NH}$ dipolar coupling restraints, 1.0, 0.035, 0.050 and 0.108 kcal mol^{-1} Hz^{-2} for the normalized (relative to $^1D_{NH}$) $^1D_{CαH}$(NH), $^1D_{CαC'}$(NH), $^1D_{NC'}$(NH) and $^1D_{HNC'}$(NH) dipolar coupling restraints respectively, and 1.0 for the conformational database potential.
[2]None of the structures exhibited interproton distance violations greater than 0.5 Å, dihedral angle violations greater than 3°, or $^3J_{HN\alpha}$ coupling constant violations greater than 2 Hz. The torsion angle restraints consist of 100 φ, 97 ψ, 80 χ$_1$, 47 χ$_2$ and 10 χ$_3$ angles. Protein backbone hydrogen bonding restraints (two per hydrogen bond, r_{NH-O} = 1.5–2.8 Å, r_{N-O} = 2.4–3.5 Å) were introduced during the final stages of refinement according to standard criteria based on amide H-D exchange experiments, backbone three-bond couplings, $^{13}C\alpha/^{13}C\beta$ secondary shifts and NOE data. The distance restraints involving the single bound water comprised three interproton distance restraints between backbone amides and water and five distance restraints for three hydrogen bonds from two backbone carbonyls and one backbone amide to water. Only structurally useful intraresidue NOEs are included in the restraints (that is, involving protons separated by more than three bonds).
[3]The Lennard-Jones van der Waals energy was calculated with the CHARMM PARAM19/20 parameters and is not included in the target function for simulated annealing or restrained minimization.
[4]The overall quality of the structure was assessed using the program PROCHECK[44]. There were no φ/ψ angles in the disallowed region of the Ramachandran plot. The dihedral angle G-factors for φ/ψ, χ$_1$/χ$_2$, χ$_1$, χ$_3$/χ$_4$ are -0.37 ± 0.02, 0.44 ± 0.04, 0.04 ± 0.11 and -0.06 ± 0.13 respectively.
[5]Defined as the average r.m.s. difference (residues 1–101) between the final 40 simulated annealing structures and the mean coordinates.

ture and topology was analyzed using the program PROMOTIF[14]. Electrostatic calculations were performed with GRASP[44]. Calculation, ranking and mapping of surface hydrophobic clusters was carried out as described[20–22].

Coordinates. The coordinates of the ensemble of 40 simulated annealing structures, the restrained regularized mean structure, and the complete list of experimental NMR restraints and ^1H, ^{15}N and ^{13}C assignments have been deposited in the Brookhaven Protein Data Bank (accession codes 2EZM, 2EZN and 2EZMMR).

1. Freed, E.O. & Martin, M.A. The role of human immunodeficiency virus type I envelope glycoproteins in virus infection. *J. Biol. Chem.* **270**, 23883–23886 (1995).

2. Wilkinson, D. HIV-1: cofactors provide the entry keys. *Curr. Biol.* **6**, 1051–1053 (1996).

3. D'Souza, M. P. & Harden, V. A. Chemokines and HIV-1 second receptors - confluence of two fields generates optimism in AIDS research. *Nature Med.* **2**, 1293–1300 (1996).

4. Boyd, M. R. et al. Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob. Agents Chemother.* **41**, 1521–1530 (1997).

5. Mori, T. et al. Analysis of sequence requirements for biological activity of cyanovirin-N, a potent HIV-inactivating protein. *Biochem. Biophys. Res. Comm.* **238**, 218–222 (1997).

6. Boyd, M.R. In *AIDS, etiology, diagnosis, treatment and prevention.* (DeVita, V. R., Hellman, S. & Rosenberg, S. A., eds) 305-319 (Alan Liss, New York; 1988).

7. Gustafson, K.R. et al. Isolation, primary sequence determination, and disulfide bond structure of cyanovirin-N, an anti-HIV protein from the cyanobacterium *Nostoc ellipsosporum. Biochem. Biophys. Res. Comm.* **238**, 223–228 (1997).

8. Mori, T. et al. Recombinant production of cyanovirin-N, a potent human immunodeficiency virus-inactivating protein derived from a cultured cyanobacterium. *Protein Exp. Purific.* **12**, 151–158 (1998).

9. Clore, G.M. & Gronenborn, A.M. Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252**, 1390–1399 (1991).

10. Clore, G. M. & Gronenborn, A. M. Determining the structures of larger proteins and protein complexes by NMR. *Trends Biotech.* **16**, 22–34 (1998).

11. Bax, A. & Grzesiek, S. Methdological advances in protein NMR *Acc. Chem. Res.* **26**, 131–138 (1993).

12. Nilges, M., Gronenborn, AM., Brünger, A. T. & Clore G. M. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Prot. Engng.* **2**, 27–38 (1988).

13. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. Use of dipolar ¹H-¹⁵N and ¹H-¹³C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nature Struct. Biol.* **4**, 732–738 (1997).

14. Hutchinson, E. G. & Thornton, J. M. PROMOTIF - a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220 (1996).

15. Altschul, S. F. et al. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

16. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).

17. Kuriyan, J. & Cowburn, D. Structures of SH2 and SH3 domains. *Curr. Opin. Struct. Biol.* **3**, 828–837 (1993).

18. Edmondson, S. P., Qiu, L., Schriver, J. W. Solution structure of the DNA-binding domain of Sac7d from the hyperthermophile *Sulfolobus acidocaldarius. Biochemistry* **34**, 13289–13304 (1995).

19. Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13–20 (1996).

20. Young, L., Jernigan, R. L. & Covell, D. G. A role for surface hydrophobicity in protein- protein recognition. *Prot. Sci.* **3**, 717–729 (1994).

21. Covell, D. G., Smythers, G. W., Gronenborn, A. M. & Clore, G. M. Analysis of hydrophobicity in the α and β chemokine families and its relevance to dimerization. *Prot. Sci.* **3**, 2064–2072 (1994).

22. Villoutreix, B. O., Härdig, Y., Wallqvist, A., Covell, D. G., de Frutos, G. & Dählback, B. Structural investigation of C4p-binding protein by molecular modeling: localization of putative binding sites. *Prot. Sci.* **in the press** (1998).

23. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

24. Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. A common sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **95**, 214–220 (1991).

25. Bax, A. et al. Measurement of homo- and hetero-nuclear J couplings from quantitative J correlation. *Meth. Enz.* **239**, 79–106 (1994).

26. Ottiger, M. & Bax, A. An empirical correlation between amide deuterium isotope effects on ¹³Cα chemical shifts and protein backbone conformation. *J. Am. Chem. Soc.* **119**, 8070–8075 (1997).

27. Grzesiek, S. & Bax, A. Measurement of amide proton exchange rates and NOE with water in ¹³C/¹⁵N enriched calcineurin B. *J. Biomol. NMR* **3**, 627–638 (1993).

28. Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in dilute liquid crystalline medium. *Science* **278**, 1111–1114 (1997).

29. Ottiger, M., Delaglio, F., & Bax, A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373–378 (1998)

30. Delaglio, F., Torchia, D. A. & Bax, A. Measurement of ¹⁵N-¹³C J couplings in staphylococcal nuclease. *J. Biomol. NMR* **1**, 439–446 (1991).

31. Clore, G. M., Gronenborn, A. M. & Bax, A. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Reson.***133**, 216–221 (1998).

32. Grzesiek, S. & Bax, A. The importance of not saturating H₂O in protein NMR: application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593–12594 (1993).

33. Nilges, M. A calculational strategy for the structure determination of symmetric dimers by ¹H-NMR. *Proteins Struct. Funct. Genet.* **17**, 297–309 (1993).

34. Clore, G.M. & Gronenborn, A.M. New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5891–5898 (1998).

35. Brünger, A.T. et al. Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination. *Acta Crystallogr. D* (1998) **In the press.**

36. Garrett, D. S. et al. The impact of direct refinement against three-bond HN-CαH coupling constants on protein structure determination by NMR. *J. Magn. Reson.* **B104**, 99–103 (1994).

37. Kuszewski, J., Qin, J., Gronenborn, A. M. & Clore, G. M. The impact of direct refinement against ¹³Cα and ¹³Cβ chemical shifts on protein structure determination by NMR. *J. Magn. Reson.* **B106**, 92–96 (1995).

38. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. *J. Magn. Reson.* **B107**, 293–297 (1995).

39. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. A potential involving multiple proton chemical-shift restraints for non-stereospecifically assigned methyl and methylene protons. *J. Magn. Reson.* **B112**, 79–81 (1996).

40. Clore, G. M., Gronenborn, A. M. & Tjandra, N. Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J. Magn. Reson.* **131**, 159–162 (1998).

41. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Prot. Sci.* **5**, 1067–1080 (1996).

42. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. *J. Magn. Reson.* **125**, 171–177 (1997).

43. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14** 51-5, 29–32 (1996).

44. Nichols, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

45. Carson, M. RIBBONS 4.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).

46. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).

47. Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. & Saraste, M. Crystal structure of a Src-homology 3 (SH3) domain. *Nature* **359**, 851–855 (1992).